AD-A218 978

DTIC ACCESSION NUMBER

PHOTOGRAPH THIS SHEET

DTIC FILE COPY

LEVEL

INVENTORY

1

A Search for Factors Causing Training ...

Sept 1987

DOCUMENT IDENTIFICATION

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DISTRIBUTION STATEMENT

ACCESSION FOR
NTIS GRA&I ☒
DTIC TRAC ☐
UNANNOUNCED ☐
JUSTIFICATION

BY
DISTRIBUTION/
AVAILABILITY CODES

DISTRIBUTION	AVAILABILITY AND/OR SPECIAL	
A-1		

DISTRIBUTION STAMP

DTIC ELECTE MAR 14 1990 S E D

DATE ACCESSIONED

QUALITY INSPECTED 4

DATE RETURNED

90 03 13 080

DATE RECEIVED IN DTIC

REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A SEARCH FOR FACTORS CAUSING TRAINING COSTS TO RISE
BY EXAMINING THE U.S. NAVY'S AT, AW, AND AX
RATINGS DURING THEIR FIRST ENLISTMENT PERIOD

by

Eugene Kapua Aiu

September 1986

Thesis Advisor: Dan C. Boger

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Naval Postgraduate School

6b OFFICE SYMBOL *(If applicable)*
Code 55

7a. NAME OF MONITORING ORGANIZATION
Naval Postgraduate School

6c. ADDRESS *(City, State, and ZIP Code)*
Monterey, California 93943-5000

7b. ADDRESS *(City, State, and ZIP Code)*
Monterey, California 93943-5000

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS *(City, State, and ZIP Code)*

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO	TASK NO.	WORK UNIT ACCESSION NO.

11 TITLE *(Include Security Classification)*
A SEARCH FOR FACTORS CAUSING TRAINING COSTS TO RISE BY EXAMINING THE U.S. NAVY'S AT, AW, AND AX RATINGS DURING THEIR FIRST ENLISTMENT PERIOD

12 PERSONAL AUTHOR(S)
Aiu, Eu(K.

13a TYPE C... T
Master's ,is

13b. TIME COVERED
FROM ________ TO ________

14 DATE OF REPORT *(Year, Month, Day)*
1986 September

15 PAGE COUNT
120

16 SUPPLEMENTARY N... ON

17 COSATI CODES

FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Manpower Study, AT, AW, AX, Length of Basic Training, Amount of Specialized Training, Attrition

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Training costs have increased in the U.S. Navy. This study examines accession data to determine if the following events caused training costs to rise; length of basic training, attrition, and amount of specialized training. The examination of these issues is restricted to three enlisted ratings, AT, AW, and AX. The time frame encompasses year group's 77 through 84. On the basis of this limited study, there is no reason to associate these three variables with increased costs.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL
Dan C. Boger

22b. TELEPHONE *(Include Area Code)*
(408) 646-3228

22c OFFICE SYMBOL
Code 55Bo

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Approved for public release; distribution is unlimited.

A search for factors causing training costs to rise
by examining the U.S. Navy's AT, AW, and AX ratings
during their first enlistment period.

by

Eugene Kapua Aiu
Lieutenant, United States Navy
B.S., Marquette University, 1979

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1986

Author: Eugene Kapua Aiu
Eugene Kapua Aiu

Approved by: Dan C. Boger
Dan C. Boger, Thesis Advisor

R R Read
Robert R. Read, Second Reader

Peter Purdue
Peter Purdue, Chairman,
Department of Operations Analysis

Kneale T. Marshall
Kneale T. Marshall,
Dean of Information and Policy Sciences

ABSTRACT

Training costs have increased in the U.S. Navy. This study examines accession data to determine if the following events caused training costs to rise; length of basic training, attrition, and amount of specialized training. The examination of these issues is restricted to three enlisted ratings, AT, AW, and AX. The time frame encompasses year group's 77 through 84. On the basis of this limited study, there is no reason to associate these three variables with increased costs.

Thesis
P2306
c.2

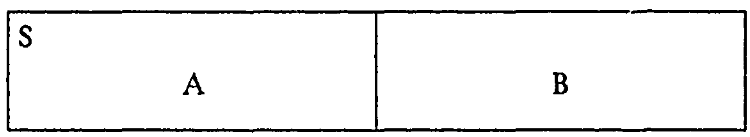
TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The U.S. Navy spends over 2 billion dollars a year on training. Training costs are rising, but the Navy does not have a clear understanding of why. A multitude of factors affect cost, however, we do *not* know what those factors are. To understand this problem, let us develop a general concept to work from. (See Figure 1.1.)



S = the set of all events that have the power to affect training costs

A = {events that have occured} ← factors

B = {events that have not occured}

$S = (A \cup B)$ $\emptyset = (A \cap B)$

Figure 1.1 The concept.

Let us identify events that have the power to affect training costs. We will call this set S. Secondly, let us divide the set S into two mutually exclusive sets A and B. Let A be the set of all events that have occured and B be the set of all events that have not occured. Our goal is to find events that belong to set A. Set A will be labeled *factors* since by definition, a factor is a contributing element that brings about a given result. In our case, the end result is rising training costs.

A. PROBLEM STATEMENT

Why is the cost of training rising? To answer this question, we divided the problem into several subproblems. We selected three subproblems to be research questions for this study.

- ***Has the length of basic training increased?***
- ***Has attrition increased?***
- ***Has the amount of specialized training increased?***

Our goal is to identify events that affect training costs. Imbedded within our problem statment are three events. These events are:

A. The length of basic training has increased.

B. Attrition has increased.

C. The amount of specialized training has increased.

Can we classify any of these events as *factors*? Or stated differently, "Have any of these events occured?" If event A, B, or C occured, then at least one reason will exist to explain the rise in training cost.

B. OBJECTIVES

This study attempts to answer three questions. Let us transform those questions into statistical hypotheses.

H_0: The length of basic training not has increased. H_1: The length of basic training has increased.
H_0: Attrition has not increased. H_1: Attrition has increased.
H_0: The amount of specialized training has not increased. H_1: The amount of specialized training has increased.

These three hypotheses form the basis of this study. Statistical methods will answer these questions by either accepting or rejecting the null hypothesis. The objectives of this thesis are:

1. Test all three hypotheses.
2. Accept or reject each event as a factor that increases cost.

II. HISTORY AND BACKGROUND

The Chief of Naval Operartions (CNO) expected training costs to fall when retention increased in the early 1980's. However, a decrease did not occur. The Center for Naval Analyses (CNA) was tasked to examine the relationships between training costs and retention. CNA formulated some general reasons why training costs might change. They set out to confirm those reasons by using information stored in their historical data files. From those data files, they provided a small data base for this study.

A. DATA BASE DESCRIPTION

The Navy has 101 enlisted rating codes. CNA's data set contains information on every enlisted rating. The data base used for this study contains information on only three enlisted ratings. These ratings are:

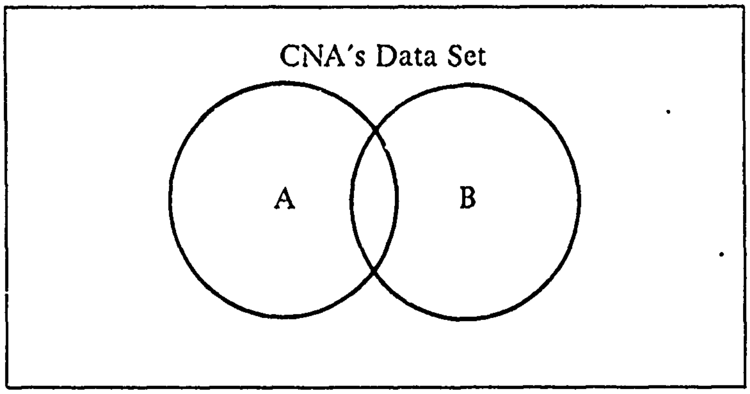
AT = Aviation Technician

AW = Aviation Anti-Submarine Warfare Operator

AX = Aviation Anti-Submarine Warfare Technician

We selected these ratings for the following reasons. This author, in conjuction with CNA, expressed an interest to examine the aviation community. Next, we decided to observe two closely related technical ratings from a squadron's maintenance department, so we selected the AT's and AX's. Lastly, we wanted to observe a rating from the squadron's operations department, so we selected the AW's.

The second point that characterizes this data base is that it is a *selected* sample from the three ratings. Given the record has a rating code of 'AT', 'AW', or 'AX', the second screening criteria consists of all records that are coded 'SG = School Guarantee'. We will say more about this criteria in the next section. Figure 2.1 provides a Venn diagram concerning the selection process for records that entered this study's data base. Corliss [Ref. 1] describes the original data set. See Appendix B for a detailed layout of this data base.



$A = (AT \cup AW \cup AX)$ $\qquad$ $B = (SG)$

Data Base $= (A \cap B)$

Figure 2.1 Record selection process.

B. EXPECTED TRAINING PATH

For the first enlistment period, an individual's expected career path follows that which is portrayed in Figure 2.2. An individual receives indoctrination at Recruit Training Command (RTC). This command is commonly known as Boot Camp. The recruit proceeds to A-school upon completion of Boot Camp. A-school provides the recruit initial skills. Upon completion of A-school, the individual advances to the fleet. The individual will receive more school based training from C-schools and F-schools, while serving productively in the fleet. C-schools and F-schools provide an individual with advanced skills and fleet skills respectively.

Let us return back to the data base selection criteria. A 'School Guarantee' is a clause written in the recruit's enlistment contract that assures the recruit will proceed directly to A-school upon completion of Boot Camp. Without the 'School Guarantee', a recruit may be sent directly to the fleet from Boot Camp. This study is strictly concerned with individuals who follow the expected training pipeline as depicted in Figure 2.2.

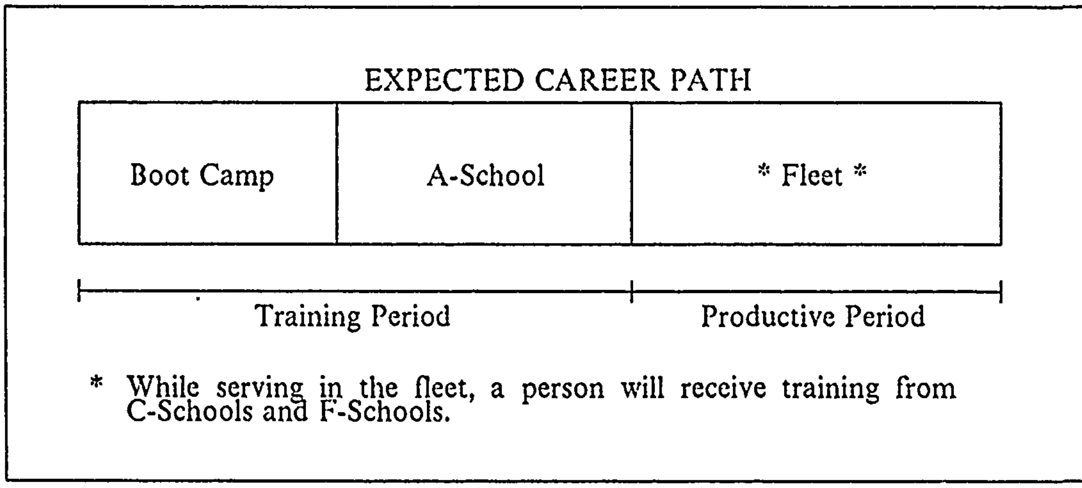


Figure 2.2 First-term enlistment milestones.

C. LIMITATIONS

As discussed earlier, the Navy has 101 enlisted ratings. However, the data base used to support this study has only three enlisted ratings. Secondly, these individuals are selected, not random. Thirdly, we are observing the performance of each group over time. The time frame is dependent upon the rating we are observing. The time frames available for study are:

AT					81	82	83	84
AW	77	78	79	80	81	82	83	84
AX					81	82	83	84

The reason for the differences in time frames is due to the fact that prior to 1981, school guarantees were not given out to individuals desiring the AT or AX ratings.

D. SCOPE

The scope of this study is restricted to the first enlistment period. (See Figure 2.3.) The following subsections describe the measures used in the analysis. Limitations and definitions are listed to set the foundation for each hypothesis test.

1. Length of Basic Training

The data base does not provide us with a way to calculate the exact time a person spends in basic training, however we have another measure. This measure is called *'time to get rated'*. (See Figure 2.4.) For each individual, we have two dates. These dates are defined as follows:

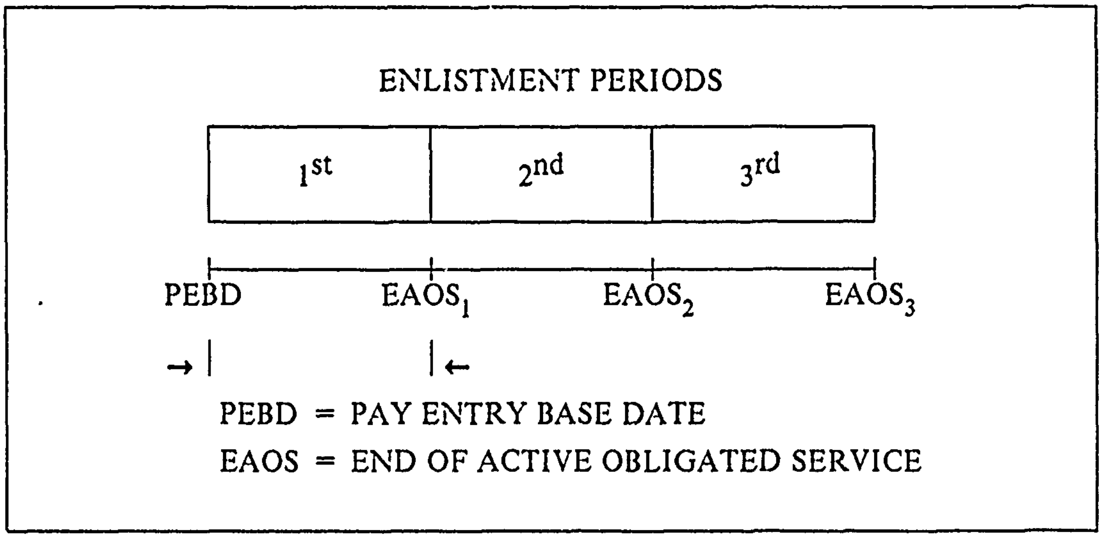


Figure 2.3 Enlistment Periods.

- PEBD = (Pay Entry Base Date) This is the date a person enters the Navy. This date is used for accounting purposes.
- RD = (Rating Date) This is the date a person is designated into one of the Navy's occupational specialties.

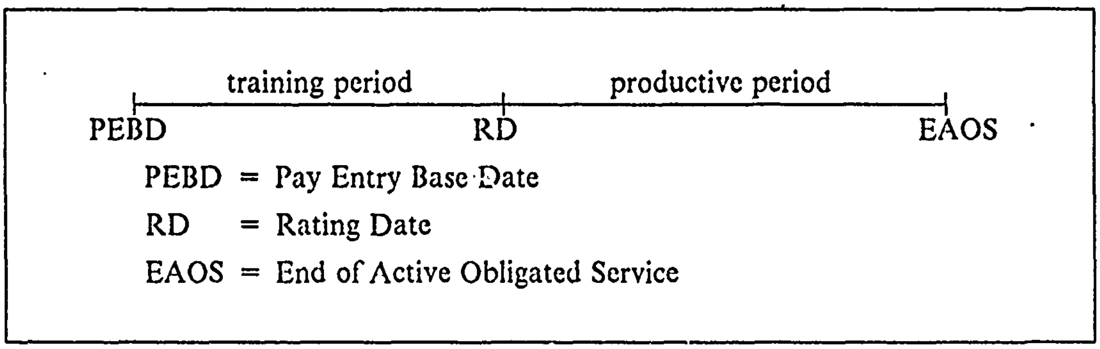


Figure 2.4 Initial Training Period.

A person gets rated upon completion of A-School or shortly thereafter. As seen in Figure 2.4, time to get rated is defined as the difference between a person's rating date and pay entry base date. Time to get rated will be used to measure the length of basic training.

As outlined in Figure 2.3, this study is restricted to the first enlistment period. This time frame is normally 48 months. The first half of the enlistment period is defined as the Basic Training period. Using this definition, our study of basic training will be restricted to the first 24 months of the enlistment period. (See Figure 2.5.)

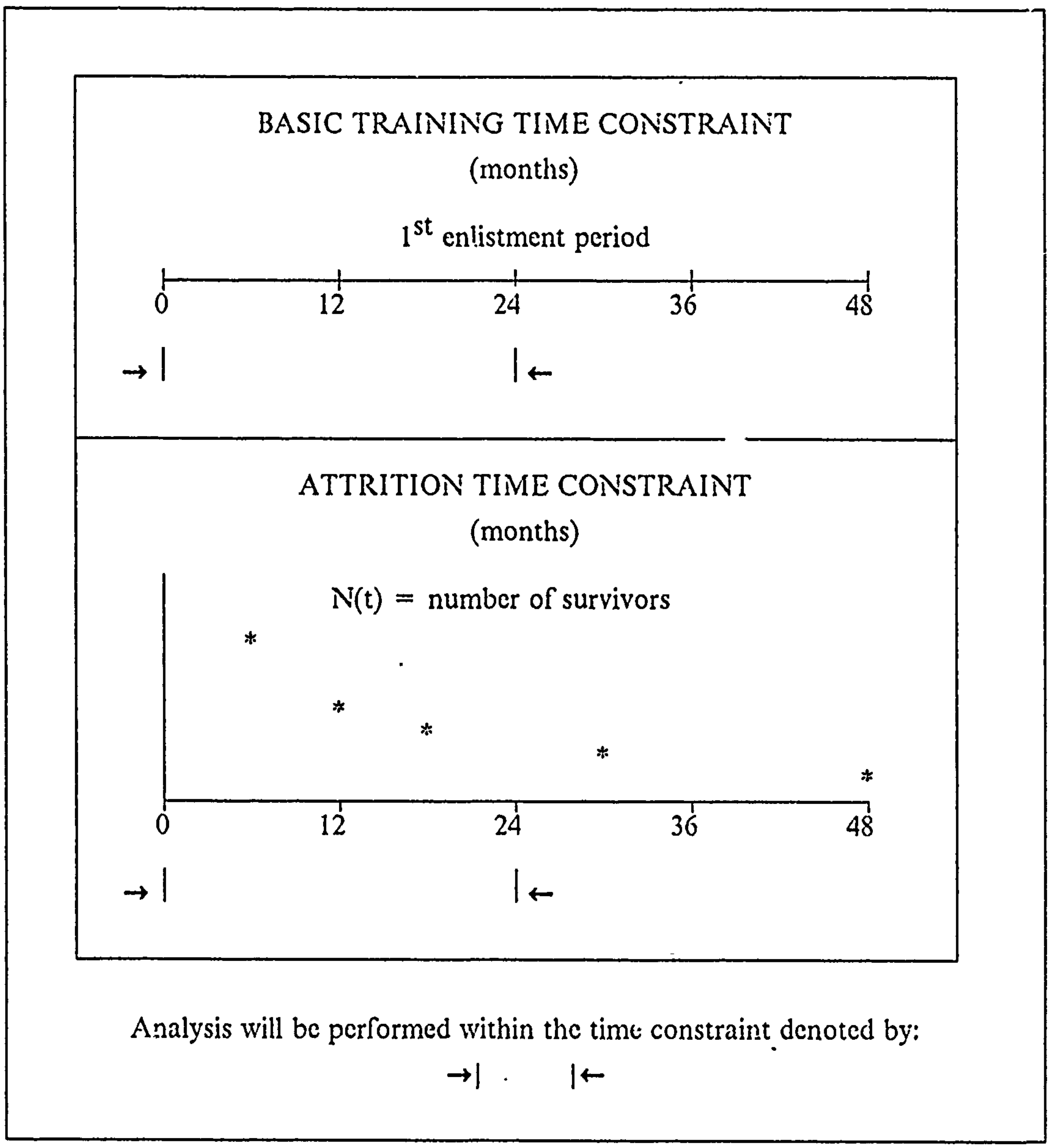


Figure 2.5 Time constraint.

2. **Attrition**

Percent losses and attrition rates are the measures used to compare year groups. Given a year group, percent loss is defined as the number of individuals that leave the Navy divided by the number of individuals that enlisted in the Navy. Attrition rate is defined as the number of individuals that leave the Navy per month. We restrict our analysis to the first 24 months per year group. Our goal is to measure attrition in the training environment and not in the operational environment. (See Figure 2.5.)

3. **Amount of specialized training**

The Navy's C-schools provide individuals with advanced/specialized skills. Upon completion of a C-school course, the individual receives a Naval Enlisted Classification (NEC) code. NEC codes supplement the enlisted rating structure by identifying particular skills in more detail than the occupational or rating structure. The navai terminology is simply this:

- **RATING = individual's occupational specialty**
- **NEC = individual's occupational subspecialty**

As an example, see Table I. Joe Sailor's occupational specialty is Aviation Technician. Joe Sailor's occupational subspecialty is:[1]

- Aircraft Radar Altimeter IMA Technician
- Aircraft Doppler Radar IMA Technician
- Aircraft Navigation Computers IMA Technician

In general, his occupation deals with aircraft navigation systems.

We measured the amount of specialized training a year group received by the number of NECs received. This measurement took place during the second and third year of service. (See Figure 2.6.)

The reasons we defined the second and third year of service as the window for analysis are threefold. One, if an individual follows the expected training pipeline, the first year is spent in Boot camp and A-school. Since the individual is not enrolled in C-school during the first year, the expected number of NEC's earned will be zero. Two, if we use the entire time period spanned by the data base, year group 78 will have had more time to aquire NEC codes than year group 80. We need to ensure each year

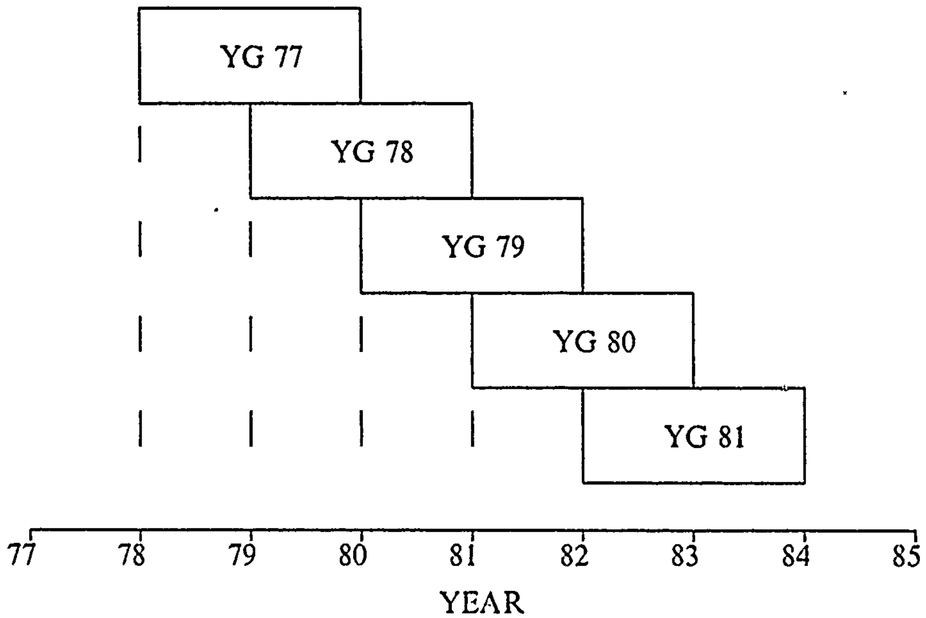
[1]The Naval Aviation Maintenance Program has three levels of maintenance. The levels are operational, intermediate, and depot. IMA is known as intermediate level maintenance.

TABLE I

JOE SAILOR'S RATING AND NEC CODES

RATING	DESCRIPTION
AT	Aviation Technician

NEC	DESCRIPTION
6605	Aircraft Radar Altimeter IMA Technician
6606	Aircraft Doppler Radar IMA Technician
6608	Aircraft Navigation Computer IMA Technician



Boxes represent the time frame a year group will be under examination

Figure 2.6 NEC Analysis Time Frames.

group has exactly the same time length and the same time period in their respective careers to accumulate NEC codes. Three, we stated earlier that our analysis will be restricted to the first enlistment period.

III. METHODOLOGY AND ANALYSIS

A. BASIC TRAINING

1. Time to get rated: Is there a trend?

Has the time to get rated changed over time? To answer this question, we define the Two Factor Analysis of Variance (ANOVA) model as follows:

	78	79	80	81	82	83	84
AT							
AW				Y_{ijk}			
AX							

MODEL: $Y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \varepsilon_{ijk}$

INDICES: i = rating
j = year group
k = k^{th} individual from group (i,j)

Y_{ijk} = number of months the k^{th} individual from group (i,j) took to get rated

μ = overall average time to get rated (grand mean)

β_i = additional time it takes an individual from rating i to get rated

τ_j = additional time it takes an individual from year group j to get rated

$(\beta\tau)_{ij}$ = interaction term

ε_{ijk} = error terms that are iid $N(0,\sigma^2)$

The goal is to test the τ vector. Is the mean time to get rated from one year group statistically different from another? We answer this question by using a statistical test. The hypothesis test and decision rule are listed in Table II.

TABLE II

TWO FACTOR ANOVA HYPOTHESIS TEST

H_0: $\tau_{77} = \ldots = \tau_{84}$
H_1: $\tau_{77} \neq \ldots \neq \tau_{84}$

H_0: The mean time to get rated has remained constant.
H_1: Not all the means are equal.

If $F^* \leq F(.95, 7, 2690)$ then conclude H_0
If $F^* > F(.95, 7, 2690)$ then conclude H_1

The other terms in the model, μ, β, and $(\beta\tau)$, are considered nuisance factors. Our goal is to account for their effects and block out their contribution. This prevents the estimate of σ^2 from being inflated. The main goal is to test for differences among year groups.

Table III lists the results of the test. All main factors are significant. Look at the table results concerning the τ vector. It is statistically significant at the .0001 level. It is highly unlikely that the τ's are equal. The P value (.0001) supports the alternate hypothesis, not all the means are the equal. Using our decision rule, since $F^* > F$, we accept the alternate hypothesis and conclude a trend exists. *"The time to get rated has changed over the years."*

Figure 3.1 is a scatter plot of the entire population. A couple of interesting things are worth noting.

- Outliers are located above the mean, none below.
- On the average, Year Group 84 took the least amount of time to get rated.
- The dispersion about the population means is smallest within Year Group 84.

Notice the presence of outliers on the high side but none on the low side. As expected, there is some minimum time required to get rated but no upper bound. We will truncate all values of Y greater than 24 months in the ensuing analysis. The reasons are threefold. One, as stated in the original set of objectives, the focus on Basic Training will be restricted to the first two years of service. Two, a set of unusual circumstances caused these individuals to take a substantial amount of time to get rated. They have detoured from the expected training pipeline and we are not interested in these individuals. Three, truncating the outliers will stabilize the variance for future ANOVA tests. Only 25 data points will be lost. This amounts to .009 or .9% of the observations. Censoring these data points should not affect future tests.

Now, let us look at 1984. Tables IV and V display Tukey's pairwise comparisons for all year groups. All pairwise comparisons with year group 84 are statistically significant. Since the average time to get rated by Year Group 84 is least among all other year groups, we will delete that group from the ensuing analysis. No further analysis need be done to that year group.

In summary, this first test establishes a trend. The time to get rated has changed over the years. Secondly, the time to get rated has decreased from 1983 to 1984. Let us investigate what happened prior to 1984.

2. Has the time to get rated increased or decreased through 1983?

The first test revealed the presence of a trend. The test also pointed out that the time to get rated decreased from 1983 to 1984 for all groups. To see what happened prior to 1984, we will test each group separately. We will follow the methodology used in Neter, Wasserman, and Kutner [Ref. 2: Sec. 17.2]. The objectives are:

- Estimate the mean time to get rated for each year group.
- Test the means for statistical difference.
- Rank the means using a paired comparison test.

Our analytical tool to test the means for statistical differences is the Single Factor ANOVA Model. The Kruskal-Wallis (KW) nonparametric test for equal means will be used as a backup test. Then, given the means are different, Tukey's paired comparison test will be used to examine the nature of the differences. Based on the paired comparison test results, we will rank the means.

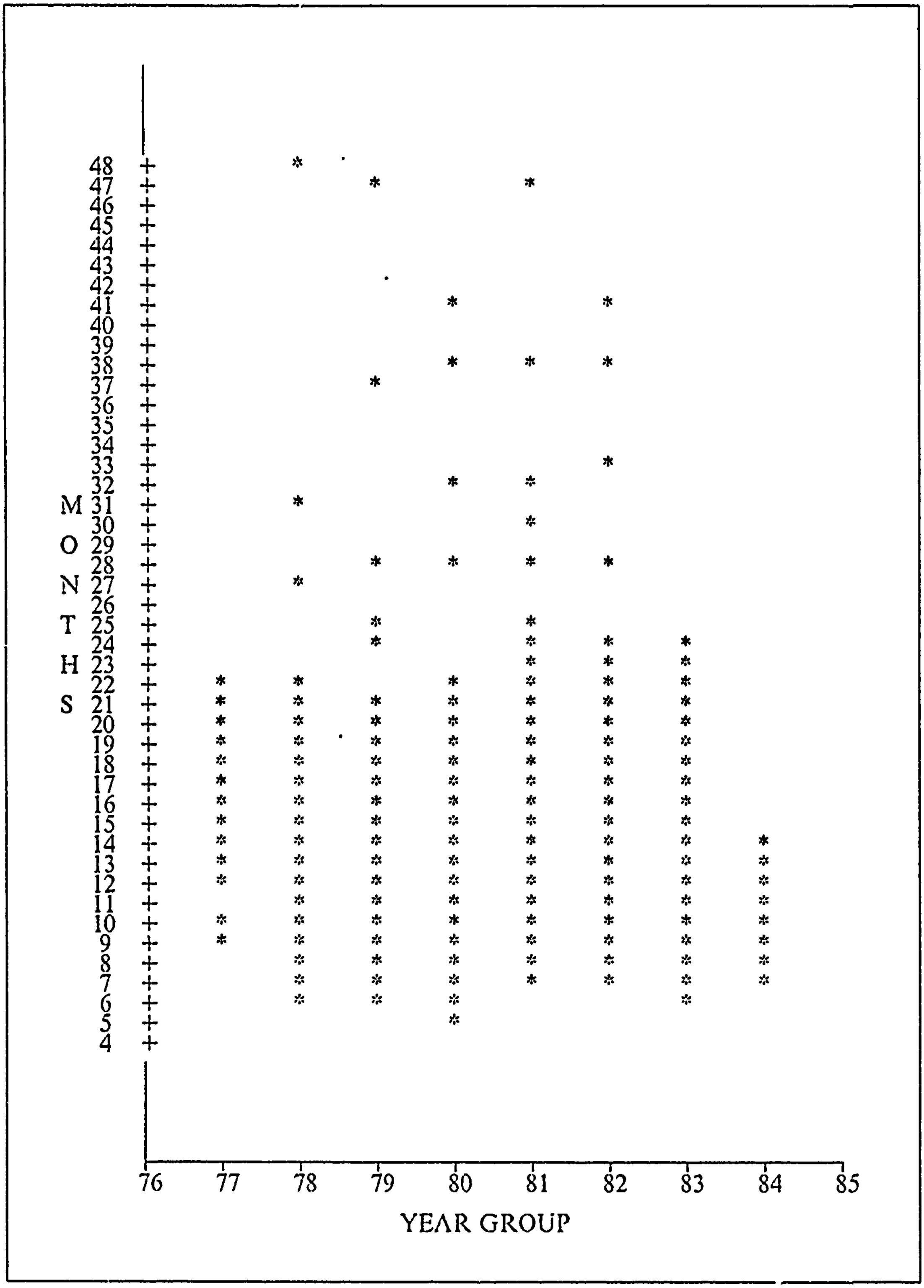


Figure 3.1 Time to get rated.

TABLE III

TIME TO GET RATED
TWO FACTOR ANOVA RESULTS

CLASS	LEVELS	VALUES
β	3	AT AW AX
τ	8	77 78 79 80 81 82 83 84

S	df	SS	MS	F*	PR > F*
Model	15	7310.0331	487.3355	34.51	0.0001
Error	2690	37984.2551	14.1205		
Total	2705	45294.2882			

S	df	SS	MS	F*	PR > F*
β	2	247.7304	123.8652	8.77	0.0002
τ	7	2844.0842	406.2977	28.77	0.0001
$\beta\tau$	6	160.8890	26.8148	1.90	0.0774

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.1614	23.1616	3.7577	16.2239

$$F(.95,7,2690) = 2.01$$

TABLE IV

TIME TO GET RATED
TUKEY'S PAIRED COMPARISON TEST RESULTS A

i	n_i	$\mu + \tau_i$
77	70	17.214
78	102	15.029
79	165	13.709
80	213	15.357
81	442	15.971
82	967	16.219
83	687	17.868
84	60	10.217

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
84-83	-9.185	-7.651	-6.116	***
84-82	-7.519	-6.003	-4.486	***
84-81	-7.322	-5.754	-4.186	***
84-80	-6.806	-5.140	-3.474	***
84-79	-5.211	-3.492	-1.774	***
84-78	-6.667	-4.813	-2.958	***
84-77	-9.003	-6.998	-4.992	***
83-84	6.116	7.651	9.185	***
83-82	1.080	1.648	2.217	***
83-81	1.202	1.897	2.592	***
83-80	1.617	2.511	3.405	***
83-79	3.170	4.158	5.147	***
83-78	1.629	2.838	4.048	***
83-77	-0.777	0.653	2.083	
82-84	4.486	6.003	7.519	***
82-83	-2.217	-1.648	-1.080	***
82-81	-0.406	0.249	0.903	
82-80	-0.000	0.862	1.725	
82-79	1.550	2.510	3.470	***
82-78	0.003	1.190	2.376	***
82-77	-2.406	-0.995	0.416	
81-84	4.186	5.754	7.322	***
81-83	-2.592	-1.897	-1.202	***
81-82	-0.903	-0.249	0.406	
81-80	-0.337	0.614	1.565	
81-79	1.222	2.261	3.301	***
81-78	-0.311	0.941	2.193	
81-77	-2.710	-1.244	0.223	

α	df	MSE
.05	2690	14.1205

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 7, 2683) = 4.290

Tukey's paired comparison confidence interval: $D \pm Ts(D)$

$D = (\mu + \tau_i) - (\mu + \tau_j)$ $T = (1/\sqrt{2})q$ $s^2(D) = [(1/n_i) + (1/n_j)]MSE$

TABLE V

TIME TO GET RATED
TUKEY'S PAIRED COMPARISON TEST RESULTS B

i	n_i	$\mu + \tau_i$
77	70	17.214
78	102	15.029
79	165	13.709
80	213	15.357
81	442	15.971
82	967	16.219
83	687	17.868
84	60	10.217

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
80-84	3.474	5.140	6.806	***
80-83	-3.405	-2.511	-1.617	***
80-82	-1.725	-0.862	0.000	
80-81	-1.565	-0.614	0.337	
80-79	0.466	1.648	2.830	***
80-78	-1.045	0.327	1.700	
80-77	-3.428	-1.857	-0.287	***
79-84	1.774	3.492	5.211	***
79-83	-5.147	-4.158	-3.170	***
79-82	-3.470	-2.510	-1.550	***
79-81	-3.301	-2.261	-1.222	***
79-80	-2.830	-1.648	-0.466	***
79-78	-2.756	-1.320	0.115	
79-77	-5.131	-3.505	-1.879	***
78-84	2.958	4.813	6.667	***
78-83	-4.048	-2.838	-1.629	***
78 82	-2.376	-1.190	-0.003	***
78-81	-2.193	-0.941	0.311	
78-80	-1.700	-0.327	1.045	
78-79	-0.115	1.320	2.756	
78-77	-3.954	-2.185	-0.416	***
77-84	4.992	6.998	9.003	***
77-83	-2.083	-0.653	0.777	
77-82	-0.416	0.995	2.406	
77-81	-0.223	1.244	2.710	
77-80	0.287	1.857	3.428	***
77-79	1.879	3.505	5.131	***
77-78	0.416	2.185	3.954	***

α	df	MSE
.05	2690	14.1205

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 7, 2683) = 4.290

Tukey's paired comparison confidence interval: $D \pm Ts(D)$

$D = (\mu + \tau_i) - (\mu + \tau_j)$ $\quad T = (1/\sqrt{2})q$ $\quad s^2(D) = [(1/n_i) + (1/n_j)]MSE$

	77	78	79	80	81	82	83
AW				Y_{ij}			

MODEL: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

INDICES: i = year group
j = j^{th} individual from year group i

Y_{ij} = number of months the j^{th} individual from rating i took to get rated

μ = overall average time to get rated

τ_i = additional time it takes an individual from year group i to get rated

ε_{ij} = error terms that are iid $N(0,\sigma^2)$

The hypothesis test and decision rules associated with the Analysis of Variance model and the Kruskal-Wallis nonparametric test are listed in Table VI.

Test results, tables, and figures that support this discussion are grouped together. They are laid out in the following manner.

AT	Figure 3.2	Data Analysis Graphs
	Table VII	ANOVA/KW test results
	Figure 3.3	Tukey's paired comparison test results
AW	Figure 3.4	Data Analysis Graphs
	Table VIII	ANOVA/KW test results
	Figure 3.5	Tukey's paired comparison test results
AX	Figure 3.8	Data Analysis Graphs
	Table IX	ANOVA/KW test results
	Figure 3.9	Tukey's paired comparison test results

Figures 3.2, 3.4, and 3.8 provide a graphical summary of the data sets. Tables VII, VIII, and IX provide the ANOVA test results and the Kruskal-Wallis test results. Figures 3.3, 3.5, and 3.9 provide Tukey's paired comparison test results. These figures display a graphical ranking of the means and a confidence interval for the difference in means. Specific results are listed in the figures and tables. We summarize our findings.

TABLE VI

SINGLE FACTOR ANOVA HYPOTHESIS TEST #1

H_0: $\tau_{77} = \ldots = \tau_{83}$
H_1: $\tau_{77} \neq \ldots \neq \tau_{83}$

H_0: The mean time to get rated has remained constant.
H_1: Not all the means are equal.

-ANOVA-
If $F^* \leq F(.95, \nu_1, \nu_2)$ then conclude H_0
If $F^* > F(.95, \nu_1, \nu_2)$ then conclude H_1

-KW-
If $\chi^2_{KW} \leq \chi^2(.95, \nu)$ then conclude H_0
If $\chi^2_{KW} > \chi^2(.95, \nu)$ then conclude H_1

- For all three ratings, the Analysis of Variance test and the Kruskal-Wallis test results were highly significant. The probability that the means are equal is almost zero. In all three cases we reject the null hypothesis and accept the alternate hypothesis. We conclude: *"The mean time to get rated has changed over the years."*
- For the AT selectees, the time to get rated is best described as no difference between year groups 81 and 82. However, year group 83 took an extra 1.5 months to get rated. There is a slight upward trend.
- For the AW selectees, the time to get rated is best described as cyclic. The mean time to get rated is highest in 1977. Over the next two years, the mean time to get rated drops to its lowest in 1979. After 1979, the trend is upwards for the next 4 years.

- For the AX selectees, the trend is U shaped. The mean time to get rated drops in 1982 and rises in 1983.

B. ATTRITION

Has attrition increased over the years? If the answer is yes, then attrition is a factor causing training costs to rise. A simple relationship exists between attrition and t aining costs. If the attrition rate is high, then the Navy must train more people to fulfill quotas. Increasing the number of people to be trained raises the training cost.

Two methods are used to answer the question. The first method uses the actual percent losses. The annual percent losses are inputs into the Cox and Stuart nonparametric test. The test determines whether an increasing trend exists. The second method uses a regression approach. Attrition rates are estimated using a nonlinear regression model. These rates serve as inputs into a simple linear regression model.

1. Percent Losses: Is it rising?

	78	79	80	81	82	83	84
Boot Camp				P_{ij}			
A-school							

P_{ij} = percent loss from school i and year group j

= number of individuals that left the Navy from group (i,j) divided by the number of individuals that started in group (i,j)

Percent losses were calculated twice, once for Boot Camp and once for A-school. We examined the sequence of numbers for an upward trend by using the Cox and Stuart nonparametric test. Conover [Ref. 3: p. 133] outlines the test procedures in detail.

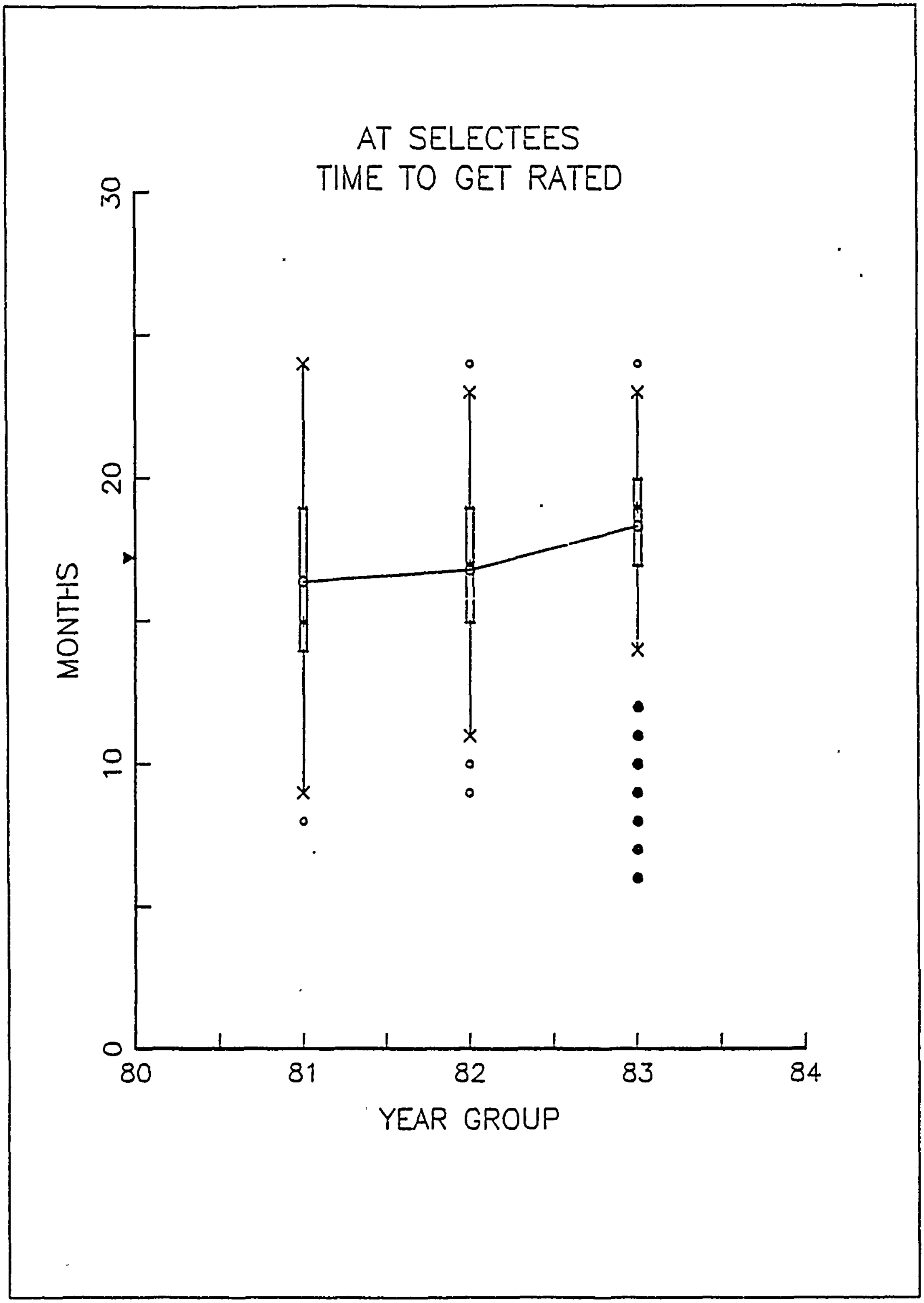


Figure 3.2 AT: Time to get rated.

TABLE VII

AT: TIME TO GET RATED
ANOVA TEST RESULTS

i	n_i	%	$\mu + \tau_i$	σ_i	-PERCENTILES- 0.25	0.50	0.75
81	226	20.3	16.385	3.5600	14	15	19
82	521	46.9	16.785	2.7153	15	17	19
83	365	32.8	18.321	3.1104	17	19	20
	1112	100.0	17.208	3.1130	15	18	19

CLASS	LEVELS	VALUES
τ	3	81 82 83

S	df	SS	MS	F*	PR > F*
Model	2	698.0855	349.0428	37.92	0.0001
Error	1109	10206.9280	9.2037		
Total	1111	10905.0135			

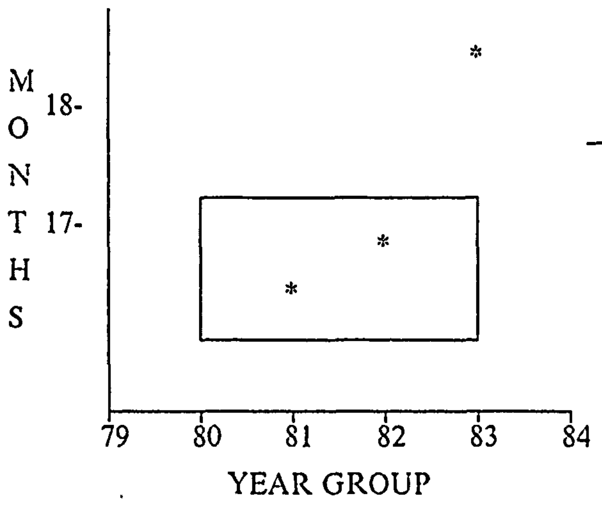
R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0640	17.6302	3.0338	17.2077

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 2)$
2	105.17	0.00

$F(.95,2,1109) = 3.00 \quad \chi^2(.95,2) = 5.99$

AT SELECTEES



i	n_i	$\mu + \tau_i$
81	226	16.385
82	521	16.785
83	365	18.321

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
83-82	1.050	1.536	2.022	***
83-81	1.333	1.936	2.538	***
82-83	-2.022	-1.536	-1.050	***
82-81	-0.167	0.400	0.967	
81-83	-2.538	-1.936	-1.333	***
81-82	-0.967	-0.400	0.167	

α	df	MSE
.05	1109	9.2037

Means boxed together are not statistically different

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 2, 1107) = 3.319

Tukey's paired comparison confidence interval: $D \pm Ts(D)$

where: $D = (\mu + \tau_i) - (\mu + \tau_j)$

$T = (1/\sqrt{2})q$

$s^2(D) = [(1/n_i) + (1/n_j)]MSE$

Figure 3.3 AT: Tukey's paired comparison test results #1.

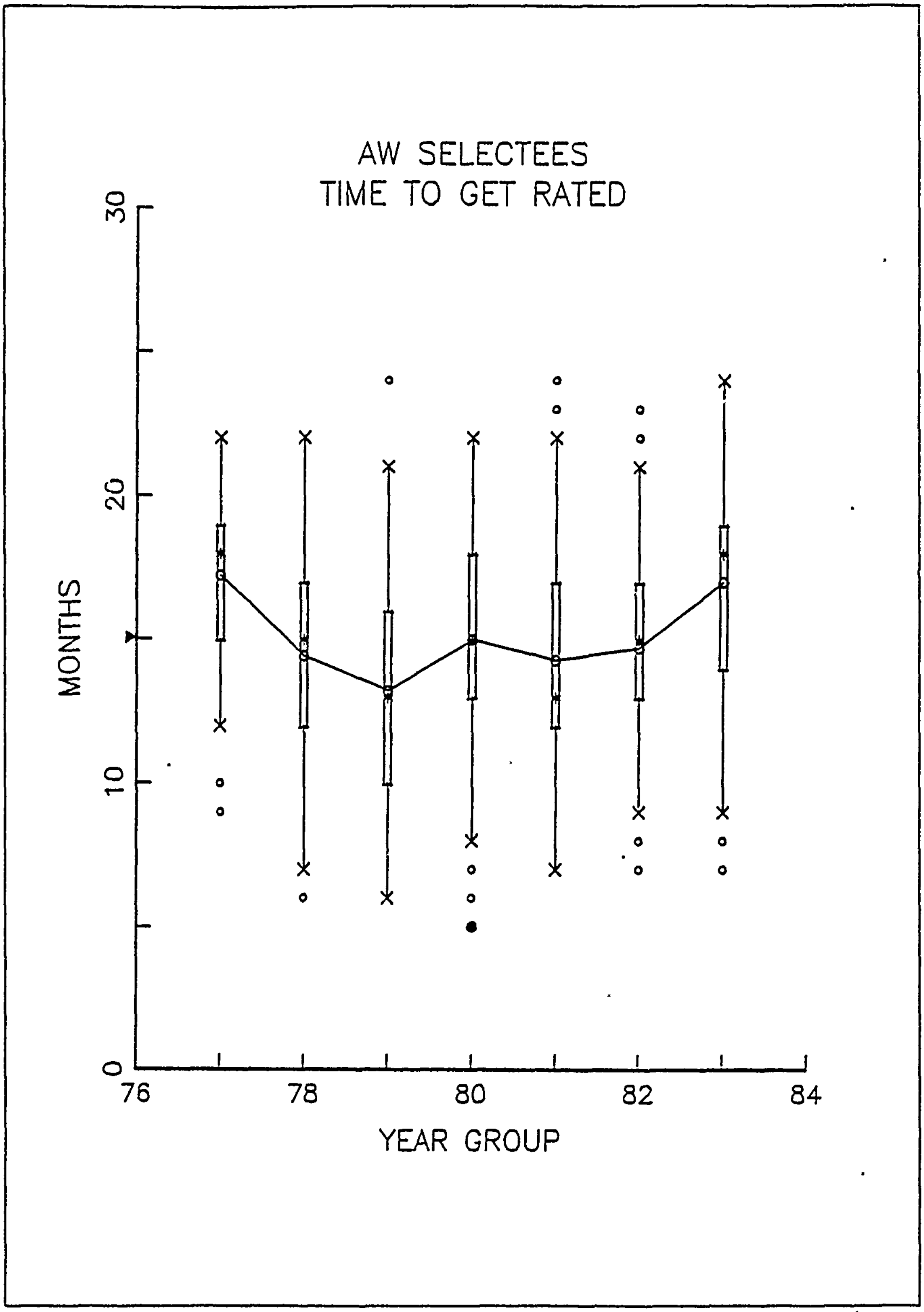


Figure 3.4 AW: Time to get rated.

TABLE VIII

AW: TIME TO GET RATED
ANOVA TEST RESULTS

i	n_i	%	$\mu + \tau_i$	σ_i	-PERCENTILES- 0.25	0.50	0.75
77	70	5.6	17.214	3.1249	15	18	19
78	99	7.9	14.414	3.7743	12	15	17
79	161	12.8	13.199	3.7895	10	13	16
80	209	16.6	14.986	3.9559	13	15	18
81	174	13.8	14.270	3.8829	12	13	1
82	303	24.0	14.703	3.2781	13	15	17
83	243	19.3	16.988	3.2721	14	18	19
	1259	100.0	15.056	3.3781	12	15	18

CLASS	LEVELS	VALUES
τ	7	77 78 79 80 81 82 83

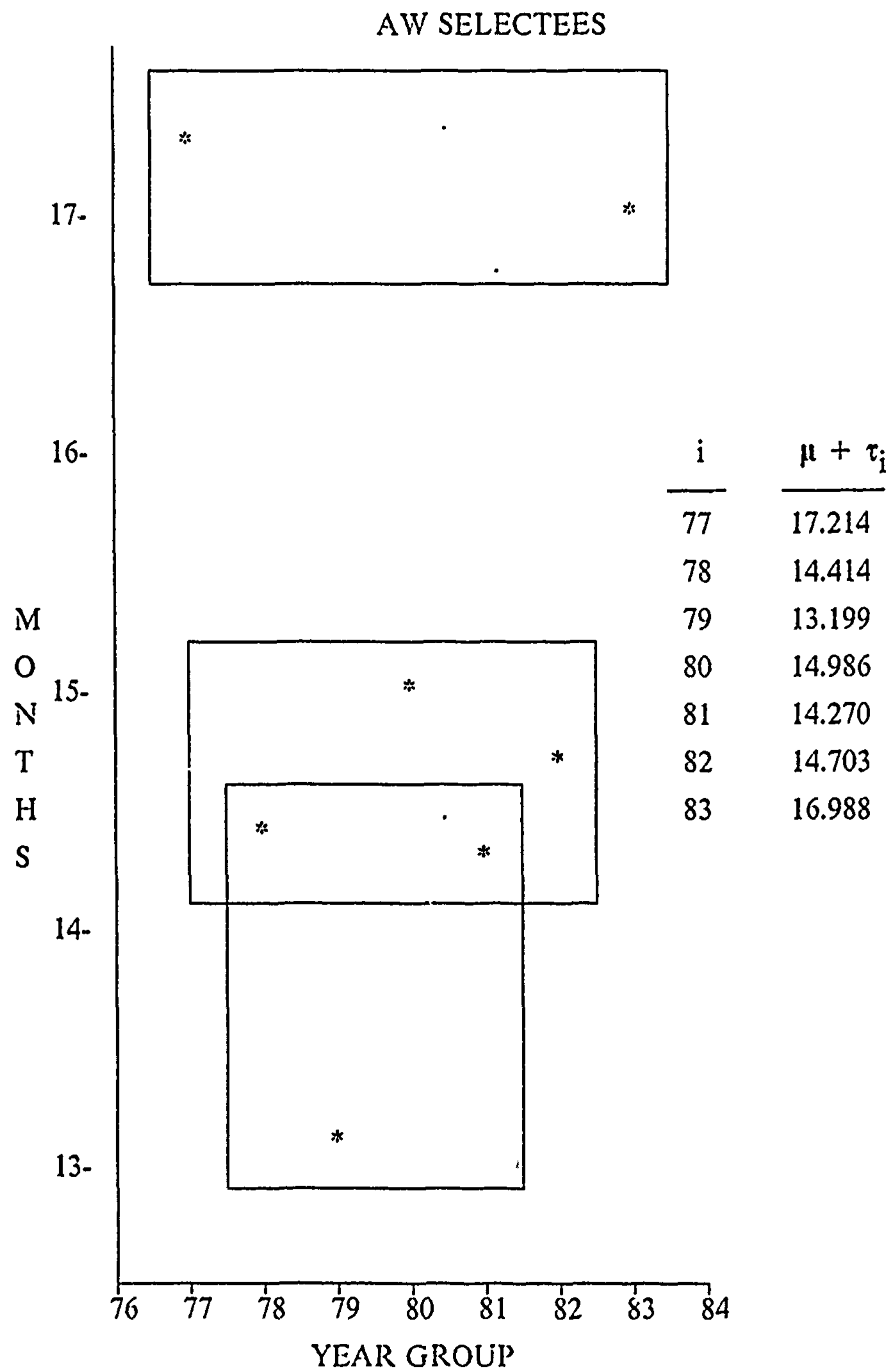
S	df	SS	MS	F*	PR> F*
Model	6	1975.1705	329.1951	25.65	0.0001
Error	1252	16066.9375	12.8330		
Total	1258	18042.1080			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.1095	23.7939	3.5823	15.0556

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 6)$
6	144.43	0.00

$F(.95,6,1252) = 2.10 \quad \chi^2(.95,6) = 12.59$



Means boxed together are not statistically different.

Figure 3.5 AW: Tukey's paired comparison test results #1a.

AW SELECTEES

i	n_i	$\mu + \tau_i$
77	70	17.214
78	99	14.414
79	161	13.199
80	209	14.986
81	174	14.270
82	303	14.703
83	243	16.988

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
83-82	1.374	2.285	3.196	***
83-81	1.667	2.718	3.768	***
83-80	1.004	2.002	3.000	***
83-79	2.714	3.789	4.864	***
83-78	1.312	2.574	3.835	***
83-77	-1.662	-0.227	1.208	
82-83	-3.196	-2.285	-1.374	***
82-81	-0.573	0.433	1.439	
82-80	-1.234	-0.283	0.669	
82-79	0.472	1.504	2.536	***
82-78	-0.936	0.289	1.514	
82-77	-3.914	-2.511	-1.108	***
81-83	-3.768	-2.718	-1.667	***
81-82	-1.439	-0.433	0.573	
81-80	-1.801	-0.716	0.370	
81-79	-0.086	1.071	2.228	
81-78	-1.476	-0.144	1.188	
81-77	-4.442	-2.944	-1.447	***
80-83	-3.000	-2.002	-1.004	***
80-82	-0.669	0.283	1.234	
80-81	-0.370	0.716	1.801	
80-79	0.678	1.787	2.896	***
80-78	-0.719	0.572	1.862	
80-77	-3.690	-2.229	-0.768	***

α	df	MSE
.05	1252	12.833

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 7, 1245) = 4.176

Tukey's paired comparison confidence interval: D $\pm$ Ts(D)

where: $D = (\mu + \tau_i) - (\mu + \tau_j)$

$T = (1/\sqrt{2})q$

$s^2(D) = [(1/n_i) + (1/n_j)]MSE$

Figure 3.6 AW: Tukey's paired comparison test results #1b.

AW SELECTEES

i	n_i	$\mu + \tau_i$
77	70	17.214
78	99	14.414
79	161	13.199
80	209	14.986
81	174	14.270
82	303	14.703
83	243	16.988

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
79-83	-4.864	-3.789	-2.714	***
79-82	-2.536	-1.504	-0.472	***
79-81	-2.228	-1.071	0.086	
79-80	-2.896	-1.787	-0.678	***
79-78	-2.567	-1.215	0.136	
79-77	-5.530	-4.016	-2.501	***
78-83	-3.835	-2.574	-1.312	***
78-82	-1.514	-0.289	0.936	
78-81	-1.188	0.144	1.476	
78-80	-1.862	-0.572	0.719	
78-79	-0.136	1.215	2.567	
78-77	-4.452	-2.800	-1.148	***
77-83	-1.208	0.227	1.662	
77-82	1.108	2.511	3.914	***
77-81	1.447	2.944	4.442	***
77-80	0.768	2.229	3.690	***
77-79	2.501	4.016	5.530	***
77-78	1.148	2.800	4.452	***

α	df	MSE
.05	1252	12.833

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 7, 1245) = 4.176

Tukey's paired comparison confidence interval: $D \pm Ts(D)$

where: $D = (\mu + \tau_i) - (\mu + \tau_j)$

$T = (1/\sqrt{2})q$

$s^2(D) = [(1/n_i) + (1/n_j)]MSE$

Figure 3.7 AW: Tukey's paired comparison test results #1c.

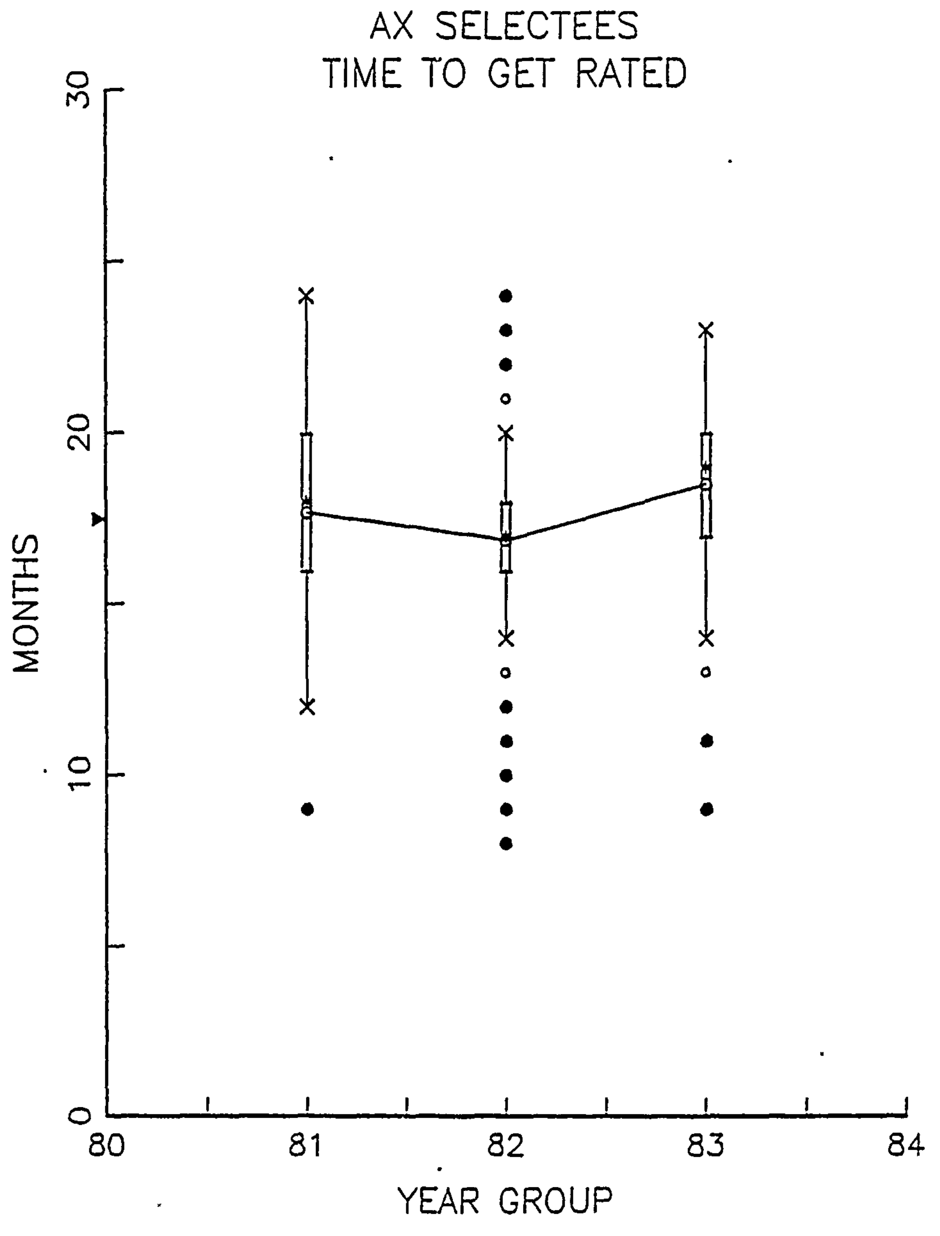


Figure 3.8 AX: Time to get rated.

TABLE IX

AX: TIME TO GET RATED
ANOVA TEST RESULTS

i	n_i	%	$\mu + \tau_i$	σ_i	-PERCENTILES- 0.25	0.50	0.75
81	33	13.1	17.667	3.4157	16	18	20
82	139	55.4	16.863	2.9073	16	17	18
83	79	31.5	18.481	2.7496	17	19	20
	251	100.0	17.478	3.0084	16	18	20

CLASS	LEVELS	VALUES
τ	3	81 82 83

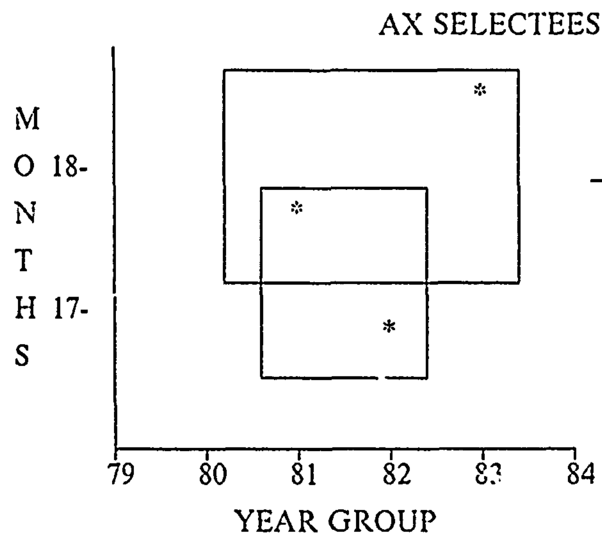
S	df	SS	MS	F*	PR> F*
Model	2	133.1718	66.5859	7.75	0.0005
Error	248	2129.4577	8.5865		
Total	250	2262.6295			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0589	16.7654	2.9303	17.4781

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 2)$
2	23.846	0.00

$F(.95,2,248) = 3.00 \quad \chi^2(.95,2) = 5.99$



i	n_i	$\mu + \tau_i$
81	33	17.667
82	139	16.863
83	79	18.481

(i, j)	CI_{lb}	$\tau_i - \tau_j$	CI_{ub}	SIG
83-82	0.644	1.618	2.591	***
83-81	-0.618	0.814	2.246	
82-83	-2.591	-1.618	-0.644	***
82-81	-2.141	-0.803	0.535	
81-83	-2.246	-0.814	0.618	
81-82	-0.535	0.803	2.141	

α	df	MSE
.05	248	8.5865

Means boxed together are not statistically different

Comparisons significant at the 0.05 level are indicated by '***'

Critical value of studentized range = q(.95; 2, 246) = 3.335

Tukey's paired comparison confidence interval: D ± Ts(D)

where: $D = (\mu + \tau_i) - (\mu + \tau_j)$

$T = (1/\sqrt{2})q$

$s^2(D) = [(1/n_i) + (1/n_j)]MSE$

Figure 3.9 AX: Tukey's paired comparison test results #1.

Cox and Stuart's test is designed to detect trends in a sequential data set. Let $X_1, \ldots, X_n$ be a sequence of random variables. The test procedures are:

1. Group the random variables into pairs $[(X_1, X_{m+1}), \ldots, (X_m, X_n)]$ where $m = n/2$.
2. Replace each pair with a (+) if $(X_{m+i} > X_i)$ or a (−) if $(X_{m+i} < X_i)$.
3. Let n equal the number of (+)'s and (−)'s. Let T^+ equal the number of (+)'s and T^- equal the number of (−)'s.
4. Set up a binomial test with parameters (n, .5).
5. Accept or reject the null hypothesis using the test statistic T^+.

Notice the arrangement of random variables. If an upward trend exists, the smallest numbers will be near the beginning of the sequence and the larger numbers near the end. The design helps to display this increasing trend. If an upward trend is present, the number of (+)'s will be greater than the number of (−)'s. If a truly random pattern existed, the number of (+)'s should be approximately equal to the number of (−)'s, $(T^+ \approx T^-)$.

To test whether the number of (+)'s is significantly different than the number of (−)'s, we use the binomial test with parameters (n,p) where $n = T^+ + T^-$ and $p = .5$.

We tested all data sets using the above procedures. Figures 3.10 through 3.21 provide the specific results. They are laid out in the following manner.

AT	Figure 3.10	Percent Losses from Boot Camp
	Figure 3.11	Cox and Stuart Test Results
	Figure 3.12	Percent Losses from A-School
	Figure 3.13	Cox and Stuart Test Results
AW	Figure 3.14	Percent Losses from Boot Camp
	Figure 3.15	Cox and Stuart Test Results
	Figure 3.16	Percent Losses from A-School
	Figure 3.17	Cox and Stuart Test Results
AX	Figure 3.18	Percent Losses from Boot Camp
	Figure 3.19	Cox and Stuart Test Results
	Figure 3.20	Percent Losses from A-School
	Figure 3.21	Cox and Stuart Test Results

Figures 3.10, 3.14, and 3.18 graphically display the percent losses from Boot Camp. Similarly, Figures 3.12, 3.16, and 3.20 graphically display the percent losses from A-school. Figures 3.11, 3.15, and 3.19 provide the Cox and Stuart test results for data sets pertaining to Boot Camp. Similarly Figures 3.13, 3.17, and 3.21 provide the Cox and Stuart test results for attrition losses in A-school. In all cases, we accepted the null hypothesis; *Attrition is not increasing.*

2. Attrition rates: Is it rising?

What is the attrition rate during basic training?
Is the attrition rate higher this year than last year?

These two questions form the basis of this subsection. Two models are presented. The first model is used to estimate the attrition rates. The second model determines if the rates are increasing.

a. Estimation of attrition rates

	78	79	80	81	82	83	84
AT							
AW				λ_{ij}			
AX							

MODEL: $N_{ij}(t) = n_{ij}e^{-\lambda_{ij}t} + \varepsilon_{ij}$

INDICES: i = rating
j = year group

$N_{ij}(t)$ = the number of survivors from group (i,j) at time t

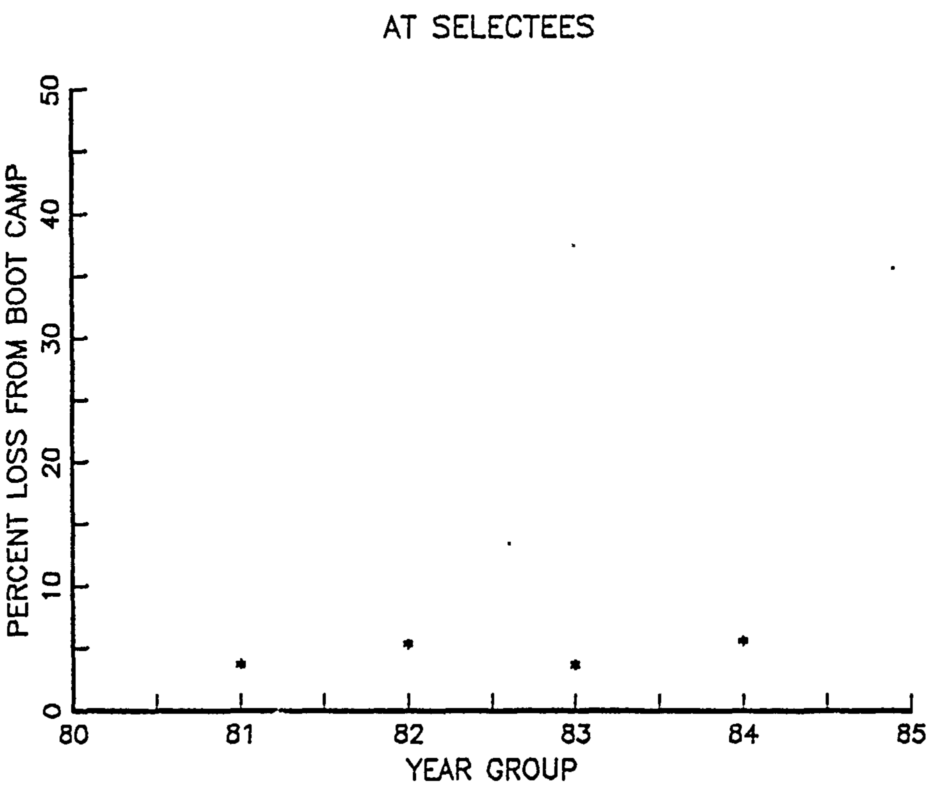
n_{ij} = the number of individuals from rating i and year group j

$e^{-\lambda_{ij}t}$ = the probability an individual from group (i,j) survives to time t

λ_{ij} = attrition rate for group (i,j)

t = time

ε_{ij} = error terms that are iid $N(0,\sigma^2)$



	81	82	83	84
STARTERS:	291	664	455	53
ATTRITES:	11	36	17	3
SURVIVORS:	280	628	438	50
LOSSES:	.0378	.0542	.0374	.0566

Figure 3.10 AT: Percent losses from Boot Camp.

AT SELECTEES

PERCENT LOSSES FROM BOOT CAMP

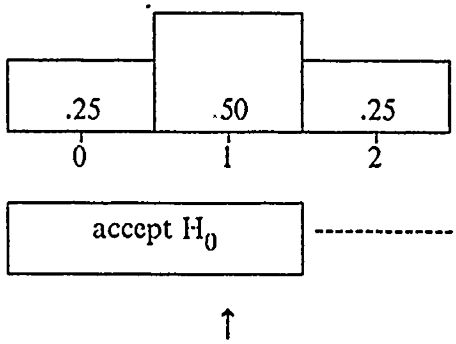
81	82	83	84
.0378	.0542	.0374	.0566

1. (.0378, .0374) (.0542, .0566)
2. $-$ $+$
3. $n = 2$ $T^{+} = 1$
4.

H_0 : Attrition is not increasing.

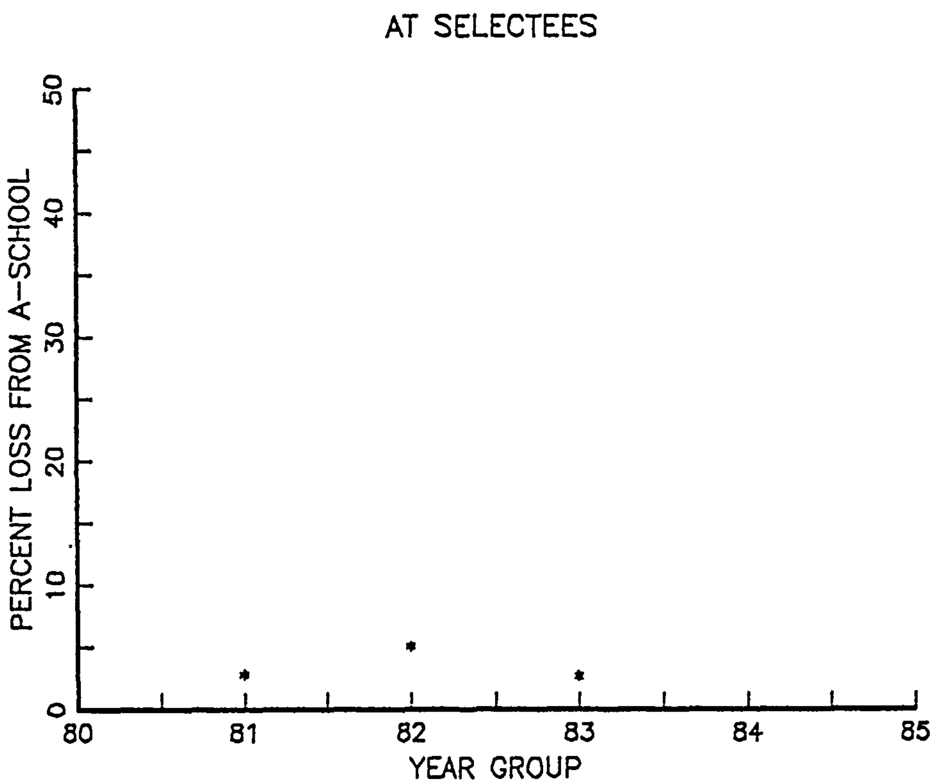
H_1 : Attrition is increasing.

$\alpha = .25$



5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.11 AT: Cox and Stuart Test Results #1.



	81	82	83	84
STARTERS:	280	628	438	50
ATTRITES:	8	32	12	0
SURVIVORS:	272	596	426	50
LOSSES:	.0286	.0510	.0274	.0000

Figure 3.12 AT: Percent losses from A-school.

AT SELECTEES

PERCENT LOSSES FROM A-SCHOOL

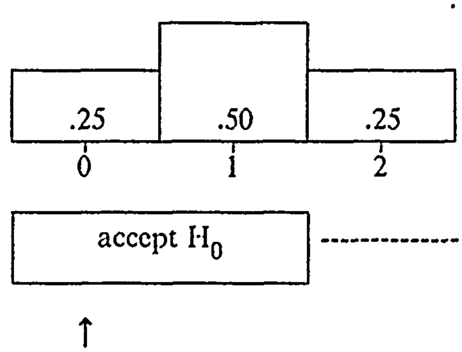
81	82	83	84
.0286	.0510	.0274	.0000

1. (.0286, .0274) (.0510, .0000)
2. – –
3. $n = 2$ $T^{+} = 0$
4.

H_0 : Attrition is not increasing.

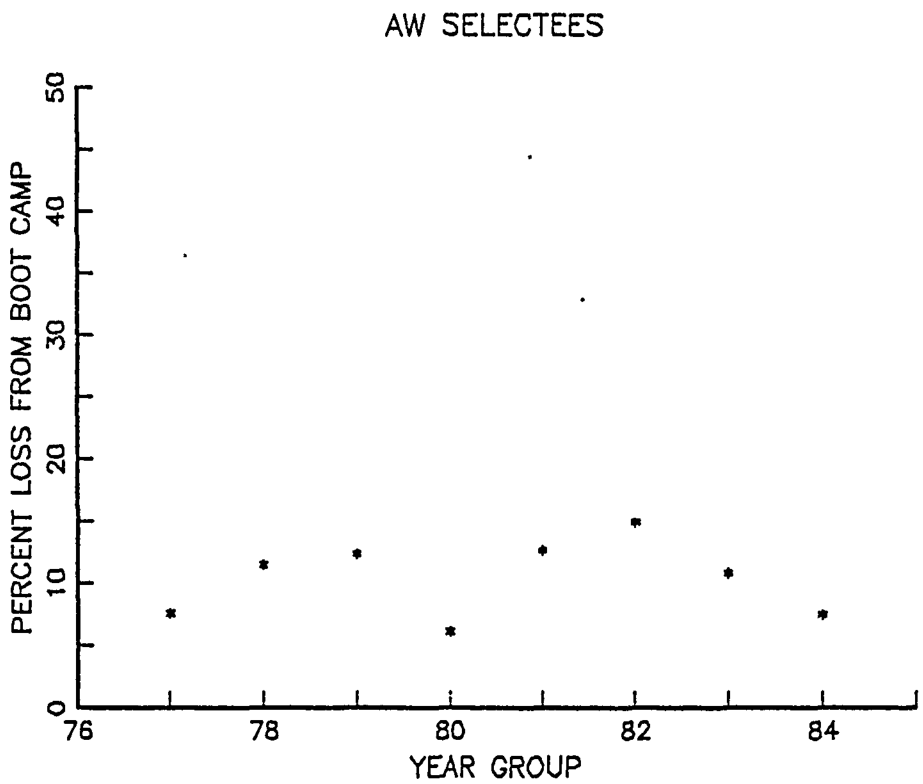
H_1 : Attrition is increasing.

$\alpha = .25$



5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.13 AT: Cox and Stuart Test Results #2.



	77	78	79	80	81	82	83	84
STARTERS:	131	173	330	324	315	501	432	53
ATTRITES:	10	20	41	20	40	75	47	4
SURVIVORS:	121	153	289	304	275	426	385	49
LOSSES:	.0763	.1156	.1242	.0617	.1270	.1497	.1088	.0755

Figure 3.14 AW: Percent losses from Boot Camp.

AW SELECTEES

PERCENT LOSSES FROM BOOT CAMP

77	78	79	80
.0763	.1156	.1242	.0617

81	82	83	84
.1270	.1497	.1088	.0755

1. (.0763, .1270) (.1242, .1088)   (.1156, .1497) (.0617, .0755)

2. $+$ $-$ $+$ $+$

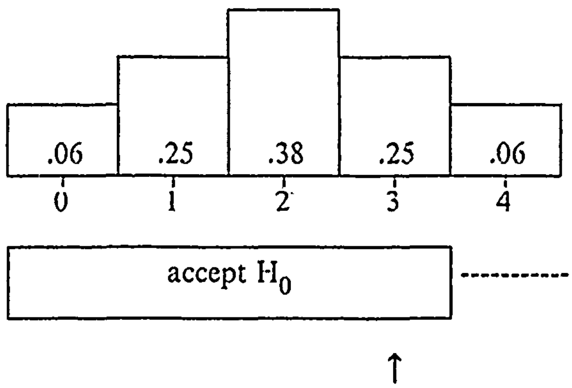
3. $n = 4$ $T^{+} = 3$

4.

H_0 : Attrition is not increasing.

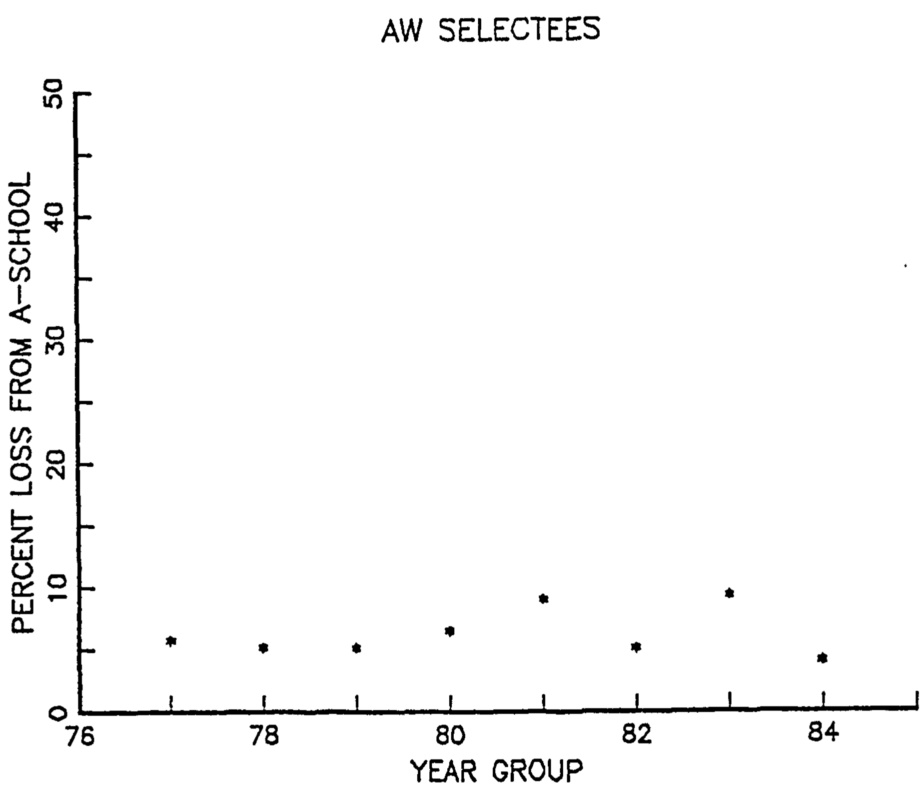
H_1 : Attrition is increasing.

$\alpha = .06$



5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.15 AW: Cox and Stuart Test Results #1.



	77	78	79	80	81	82	83	84
STARTERS:	121	153	289	304	275	426	385	49
ATTRITES:	7	8	15	20	25	22	36	2
SURVIVORS:	114	145	274	284	250	404	349	47
LOSSES:	.0579	.0523	.0519	.0658	.0909	.0516	.0935	.0408

Figure 3.16 AW: Percent losses from A-school.

AW SELECTEES

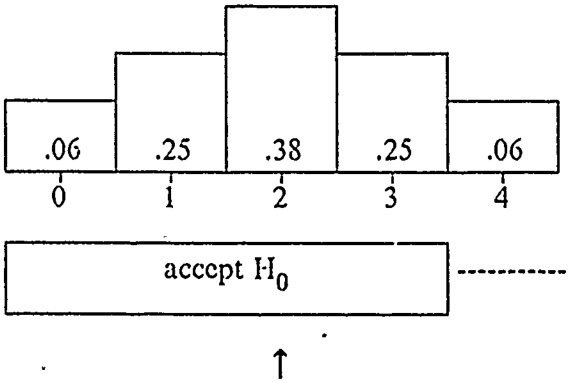
PERCENT LOSSES FROM A-SCHOOL

77	78	79	80
.0579	.0523	.0519	.0658

81	82	83	84
.0909	.0516	.0935	.0408

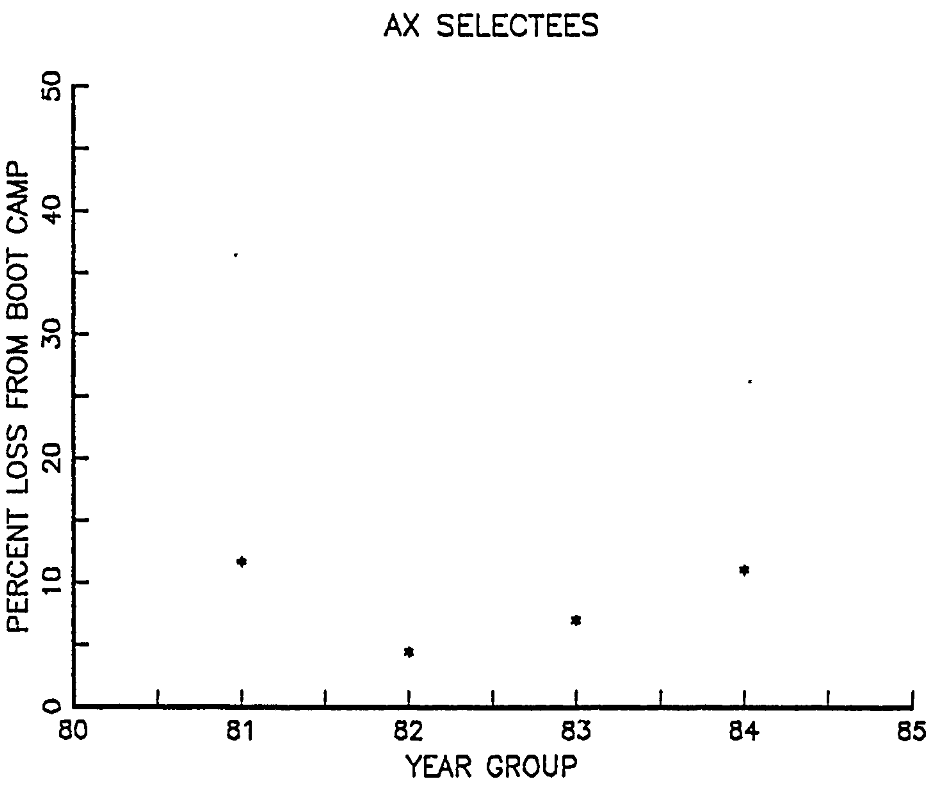
1. (.0579, .0909) (.0519, .0935)   (.0523, .0516) (.0658, .0408)
2. $+$ $+$ $-$ $-$
3. $n = 4$ $T^{+} = 2$
4. H_0 : Attrition is not increasing.
   H_1 : Attrition is increasing.

$\alpha = .06$



5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.17 AW: Cox and Stuart Test Results #2.



	81	82	83	84
STARTERS:	51	177	99	9
ATTRITES:	6	8	7	1
SURVIVORS:	45	169	92	8
LOSSES:	.1176	.0452	.0707	.1111

Figure 3.18 AX: Percent losses from Boot Camp.

AX SELECTEES

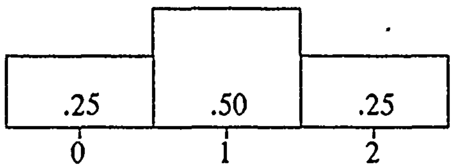
PERCENT LOSSES FROM BOOT CAMP

81	82	83	84
.1176	.0452	.0707	.1111

1. (.1176, .0707) (.0452, .1111)
2. $-$ $+$
3. $n = 2$ $T^{+} = 1$
4.

H_0 : Attrition is not increasing.
H_1 : Attrition is increasing.

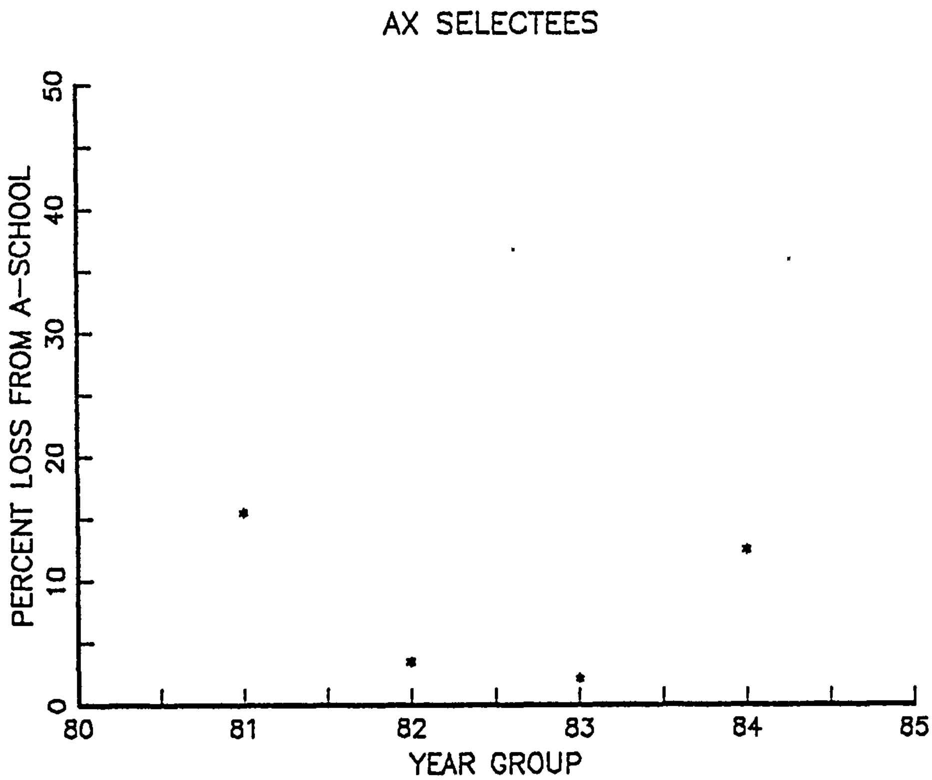
$\alpha = .25$



accept H_0 --------------

5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.19 AX: Cox and Stuart Test Results #1.



	81	82	83	84
STARTERS:	45	169	92	8
ATTRITES:	7	6	2	1
SURVIVORS:	38	163	90	7
LOSSES:	.1556	.0355	.0217	.1250

Figure 3.20 AX: Percent losses from A-school.

AX SELECTEES

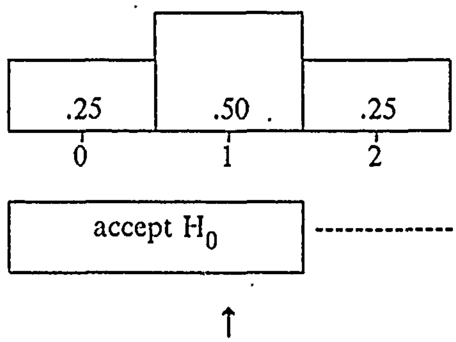
PERCENT LOSSES FROM A SCHOOL

81	82	83	84
.1556	.0355	.0217	.1250

1. (.1556, .0217) (.0355, .1250)
2. $-$ $+$
3. $n = 2$ $T^{+} = 1$
4.

H_0 : Attrition is not increasing.
H_1 : Attrition is increasing.

$\alpha = .25$



5. Since T^{+} falls in the acceptance region we accept H_0. *Attrition is not increasing.*

Figure 3.21 AX: Cox and Stuart Test Results #2.

This is a simple nonlinear model with one parameter (λ_{ij}) to be estimated per cell. Imbedded within the model is a couple of things worth mentioning. The $e^{-\lambda_{ij}t}$ term represents the probability an individual from group (i,j) remains in the Navy till time t. This is the exponential survival function. Let T_{ij} be the random variable that represents the probability distribution with survival function $e^{-\lambda_{ij}t}$. Due to the uniqueness of survival functions, $T_{ij} \sim EXP(\lambda_{ij})$. Hence, the time spent in basic training is exponentially distributed. The next term to look at is $n_{ij}e^{-\lambda_{ij}t}$. Here n_{ij} represents the number of individuals from rating i and year group j and $e^{-\lambda_{ij}t}$ is the probability an individual from group (i,j) survives till time t. So, $n_{ij}e^{-\lambda_{ij}t}$ is nothing more than the expected value of a Binomial random variable with parameters $(n, p) = (n_{ij}, e^{-\lambda_{ij}t})$. Now let's look at the model in it's entirety, [$N_{ij} = n_{ij}e^{-\lambda_{ij}t} + \varepsilon_{ij}$]. For a given t, N_{ij} can be thought of as a systematic term plus some noise (ε_{ij}). The systematic term is the expected value of a binomial distribution. It represents the expected number of survivors at time t.

Our goal is to estimate λ_{ij}. We used the NLIN procedure in SAS to estimate the parameter λ for each group. See Appendix C for a copy of the SAS program and the data vectors used by the program. Table X provides the results.

b. Are attrition rates increasing?

	77	78	79	80	81	82	83
AW				Y_i			

MODEL: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

INDEX: i = year group

Y_i = attrition rate for year group i

β_0 = constant attrition rate for all year groups

β_1 = change in Y due to a one unit change in X (slope)

X_i = year group i

ε_i = error terms that are iid $N(0,\sigma^2)$

TABLE X

ATTRITION RATES

λ = ATTRITION RATE

	AT	AW	AX
77		.0102 (.0017)	
78		.0094 (.0008)	
79		.0128 (.0012)	
80		.0087 (.0006)	
81	.0041 (.0002)	.0145 (.0015)	.0175 (.0017)
82	.0060 (.0003)	.0127 (.0013)	.0047 (.0003)
83	.0048 (.0004)	.0168 (.0013)	.0231 (.0012)
84	.0067 (.0007)	.0123 (.0024)	.0326 (.0054)

.xxxx ← estimate of λ_{ij}

(.yyyy) ← asymptotic standard error of the estimate

Recall event B defined in our problem statement: Attrtion is increasing. We will use the linear regression model [$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$] to ascertain the validity of the statement. The linear regression model permits us to statistically verify event B. We will test the regression coefficient β_1. If β_1 is statistically greater than zero, then we will conclude: "Attrition rates are increasing." Let us set up our hypothesis test.

Test number one is the F-test. As stated in Draper and Smith, [Ref. 4: p. 32], the F-test will determine if a trend exists in the regression equation. The hypothesis test and decision rule associated with this test are listed in Table XI.

TABLE XI

LINEAR REGRESSION F-TEST #1

H_0: $\beta_1 = 0$ [Attrition rates are constant.]

H_1: $\beta_1 \neq 0$ [Attrition rates are not constant.]

If $F^* \leq F(.95, 1, n-2)$ then conclude H_0

If $F^* > F(.95, 1, n-2)$ then conclude H_1

Test number two is the one sided t-test. This test is used after the F-test. If the F-test determines that a trend exists, then this test will determine the direction of the trend [Ref. 2: p. 68]. The hypothesis test and the decision rule associated with the one sided t-test are listed in Table XII.

TABLE XII

LINEAR REGRESSION t-TEST #1

H_0: $\beta_1 \leq 0$ [Attrition rates are not increasing.]

H_1: $\beta_1 > 0$ [Attrition rates are increasing.]

If $t^* \leq t(.95, n-2)$ then conclude H_0

If $t^* > t(.95, n-2)$ then conclude H_1

We performed three tests. See Figures 3.22, 3.23, and 3.24 for specific results. The F-test results are listed in Table XIII. In all three cases, $F^* \leq F$. By our decision rule, we accept H_0 and conclude: *"Attrition rates are constant."* The one sided t-test sequentially follows the F-test. Our results show that the F-test is not statistically significant. In view of this fact, it's not necessary to perform the t-test. However, details of the t-test are listed in Figures 3.22, 3.23, and 3.24. We summarize the results of the one sided t-test by saying, *"Attrition rates are not increasing."*

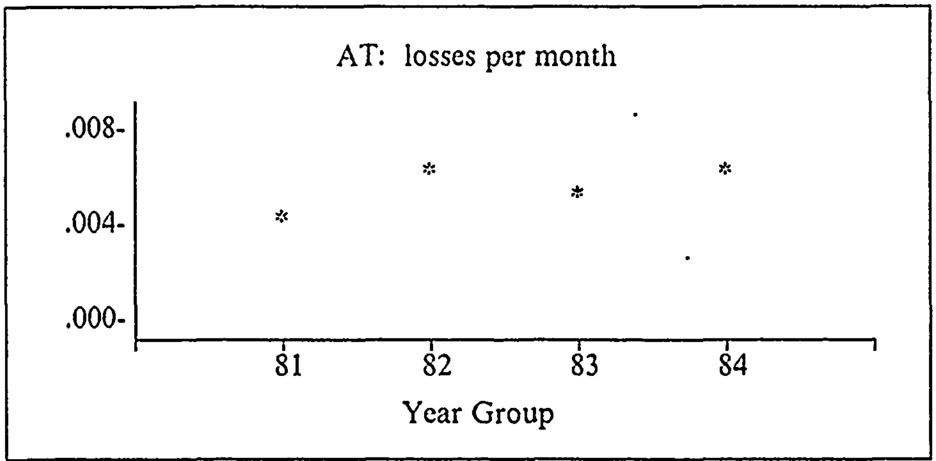
TABLE XIII

REGRESSION ON ATTRITION RATES: F-TEST RESULTS

Rating	F*	n	F(.95,1,n-2)
AT	2.14	4	18.50
AW	3.72	8	5.99
AX	1.99	4	18.50

C. SPECIALIZED TRAINING

The third event of our problem statement is: The amount of specialized training has increased. As previously discussed, we will measure the amount of specialized training by counting the number of NEC's an individual acquires. Secondly, the measurement will take place during the individual's second and third year of service. Two methods are presented. Given a year group, we looked at the average number of NEC's per individual. We plugged these numbers into a regression model and tested this sequence to determine if an increasing trend existed. Method number two used a random sample of individuals from each year group. A balanced design ANOVA model determined if the average number of NEC's per year group differed. The ensuing analysis excludes year group 84 because the data base does not cover their third year of service.



81	82	83	84
.0041	.0060	.0048	.0067

S	df	SS	MS	F*	PR> F*
Regression	1	0.000002	0.000002	2.1440	0.2807
Error	2	0.000002	0.000001		
Total	3	0.000004			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.5174	18.92083	0.001022	0.005401

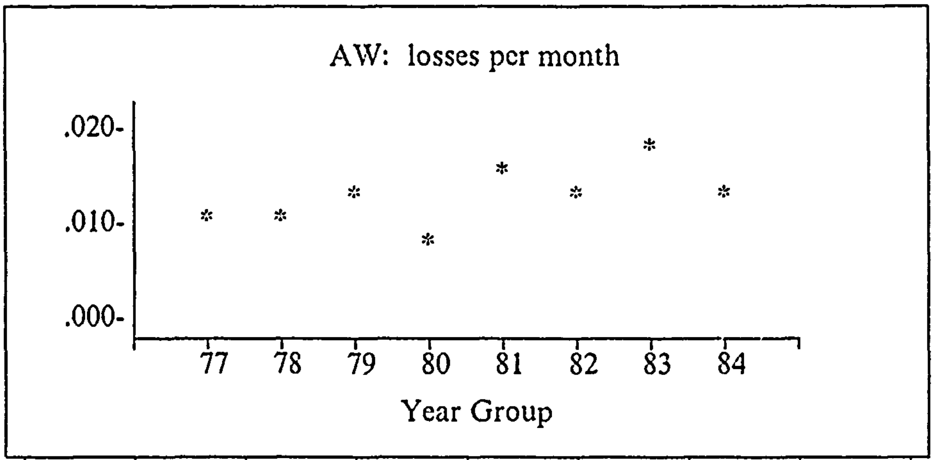
β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	0.003728	0.001252		
β_1	1	0.000669	0.000457	1.4640	0.1403

β_i	CI_{lb}	b_i	CI_{ub}
β_0	0.000072	0.003728	0.007384
β_1	-.000665	0.000457	0.002003

$F^* = MSR/MSE \quad F(.95,1,2) = 18.5 \qquad t^* = b_1/s(b_1) \quad t(.95,2) = 2.92$

CI: $b_i \pm t(.95,2)s(b_i)$

Figure 3.22 AT: Attrition rates - Regression results.



77	78	79	80	81	82	83	84
.0102	.0093	.0128	.0087	.0145	.0127	.0168	.0123

S	df	SS	MS	F*	PR> F*
Regression	1	0.000020	0.000020	3.7190	0.1021
Error	6	0.000032	0.000005		
Total	7	0.000052			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.3826	18.83448	0.002293	0.012718

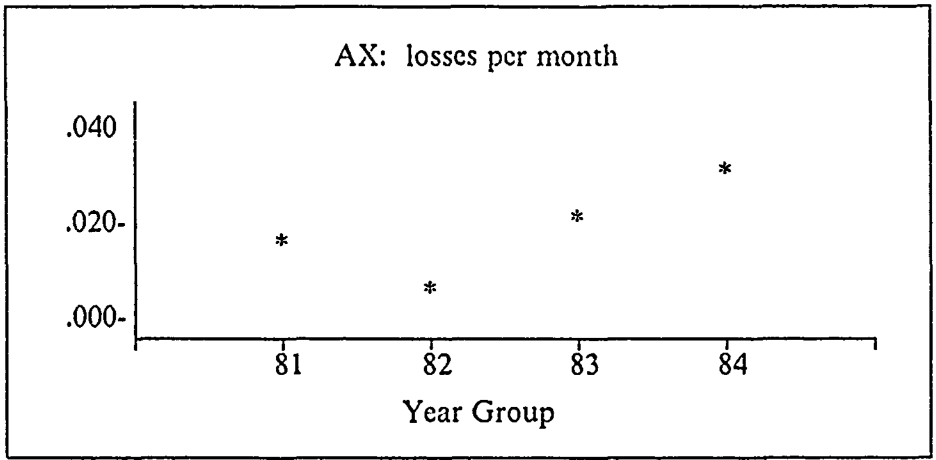
β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	0.009107	0.001787		
β_1	1	0.000683	0.000354	1.9280	0.0511

β_i	CI_{lb}	b_i	CI_{ub}
β_0	-.004729	0.009107	0.013484
β_1	-.000184	0.000683	0.001549

$F^* = MSR/MSE \quad F(.95,1,6) = 5.99 \qquad t^* = b_1/s(b_1) \quad t(.95,6) = 1.94$

CI: $b_i \pm t(.95,6)s(b_i)$

Figure 3.23 AW: Attrition rates - Regression results.



81	82	83	84
.0175	.0047	.0231	.0326

S	df	SS	MS	F*	PR> F*
Regression	1	0.000204	0.000204	1.9954	0.2933
Error	2	0.000204	0.000102		
Total	3	0.000408			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.4994	51.88053	0.010110	0.019487

β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	-.003520	0.012382		
β_1	1	0.006387	0.004521	1.4126	0.1466

β_i	CI_{lb}	b_i	CI_{ub}
β_0	-.049757	0.003520	0.056757
β_1	-.013067	0.006387	0.025841

$F^* = MSR/MSE \quad F(.95,1,2) = 18.5 \qquad t^* = b_1/s(b_1) \quad t(.95,2) = 2.92$

CI: $b_i \pm t(.95,2)s(b_i)$

Figure 3.24 AX: Attrition rates - Regression results.

1. **Average number of NEC's per individual: Has it increased?**

TABLE XIV
AVERAGE NUMBER OF NEC'S PER INDIVIDUAL

	i	NEC_i	N_i	AVG
AT	81	369	232	1.5905
	82	1010	524	1.9275
	83	619	365	1.6959
AW	77	114	70	1.6286
	78	154	102	1.5098
	79	349	165	2.1152
	80	422	213	1.9812
	81	352	177	1.9887
	82	668	304	2.1974
	83	444	243	1.8272
AX	81	58	33	1.7576
	82	255	139	1.8345
	83	133	79	1.6835

For each rating and year group, Table XIV lists the average number of NEC's per individual. This number is (NEC_i / N_i) where:

NEC_i = number of NEC's acquired by year group i

N_i = number of individuals in year group i

We will set up the regression model and statistically test these table values for an upward trend. The model is hereby defined.

	77	78	79	80	81	82	83
AW				Y_i			

MODEL: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

INDEX: i = year group

Y_i = average number of NEC's per individual from year group i

β_0 = constant number of NEC's per individual

β_1 = change in Y per unit change in X (slope)

X_i = year group i

ε_i = error terms that are iid $N(0, \sigma^2)$

The same methodology presented in the previous section will be used. The F-test will determine if a trend exists and the one sided t-test will ascertain the direction of the trend. The hypothesis tests and decision rules are presented in Tables XV and XVI.

See Figures 3.25, 3.26, and 3.27. The test results clearly show that a trend is absent. The F-test forces us to accept the null hypothesis in all three cases. Likewise, the t-test directs us to accept the null hypothesis. We conclude this subsection by saying: *"The average number of NEC's per individual is not increasing."*

2. Average number of NEC's per year group: Has it increased?

The first method for determining the amount of specialized training condensed our data base into a few observations. We all know that a small sample size does not provide a powerful statistical result. The second method uses the single factor ANOVA model. We wanted to increase the number of observations in the test and use a balanced design. We took a random sample of 30 data points from each year group and tested the sample means for statistical differences. We present the model.

TABLE XV

LINEAR REGRESSION F-TEST #2

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

H_0: The average number of NEC's per individual is constant.
H_1: The average number of NEC's per individual is not constant.

If $F^* \leq F(.95, 1, n-2)$ then conclude H_0
If $F^* > F(.95, 1, n-2)$ then conclude H_1

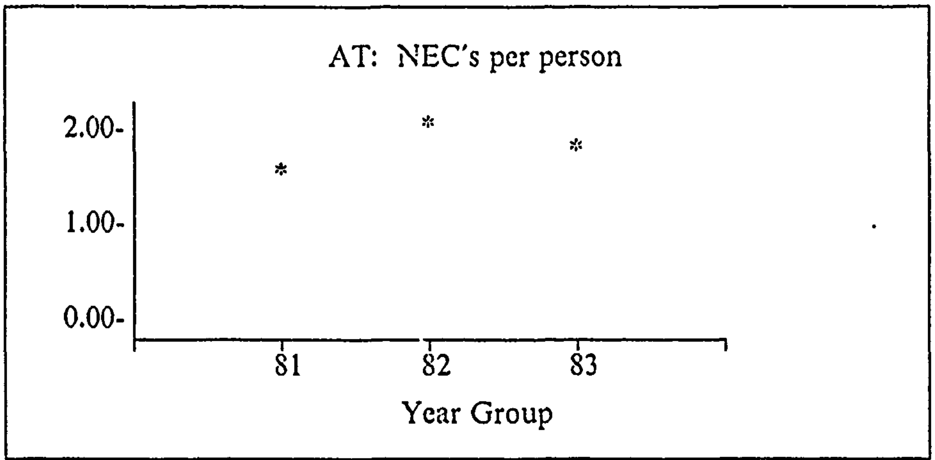
TABLE XVI

LINEAR REGRESSION t-TEST #2

H_0: $\beta_1 \leq 0$
H_1: $\beta_1 > 0$

H_0: The average number of NEC's per individual is not rising.
H_1: The average number of NEC's per individual is rising.

If $t^* \leq t(.95, n-2)$ then conclude H_0
If $t^* > t(.95, n-2)$ then conclude H_1



81	82	83
1.590	1.927	1.696

S	df	SS	MS	F*	PR> F*
Regression	1	0.005555	0.005555	0.1030	0.8022
Error	1	0.053884	0.053884		
Total	2	0.059439			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0935	13.35641	0.232130	1.737967

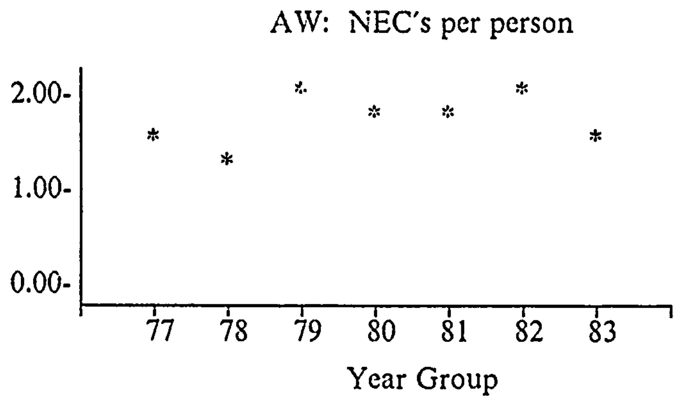
β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	1.632600	0.354580		
β_1	1	0.052700	0.164141	0.3210	0.4011

β_i	CI_{lb}	b_i	CI_{ub}
β_0	-2.81300	1.579867	6.078200
β_1	-2.00500	0.052700	2.110600

$F^* = MSR/MSE$ $F(.95,1,1) = 161$ $t^* = b_1/s(b_1)$ $t(.95,1) = 6.31$

CI: $b_i \pm t(.95,1)s(b_i)$

Figure 3.25 AT: NEC's per individual - Regression results.



77	78	79	80	81	82	83
1.629	1.510	2.115	1.981	1.989	2.197	1.827

S	df	SS	MS	F*	PR> F*
Regression	1	0.121506	0.121506	2.350	0.1859
Error	5	0.258542	0.051708		
Total	6	0.380048			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.3197	12.01502	0.227395	1.892586

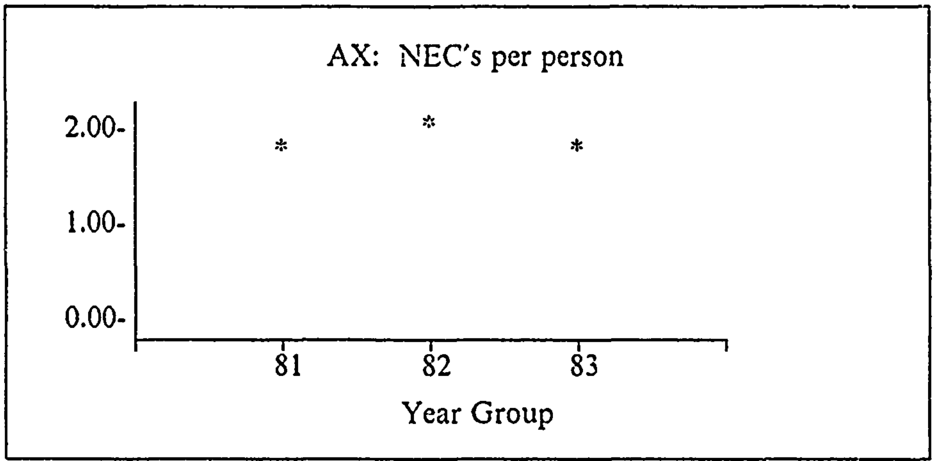
β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	1.629086	0.192184		
β_1	1	0.065875	0.042974	1.5330	0.0929

β_i	CI_{lb}	b_i	CI_{ub}
β_0	1.135100	1.629086	2.123100
β_1	-.044590	0.065875	0.176340

$F^* = MSR/MSE$ $F(.95,1,5) = 6.61$ $t^* = b_1/s(b_1)$ $t(.95,5) = 2.02$

CI: $b_i \pm t(.95,2)s(b_i)$

Figure 3.26 AW: NEC's per individual - Regression results.



81	82	83
1.758	1.834	1.683

S	df	SS	MS	F*	PR> F*
Regression	1	0.002745	0.002745	0.3170	0.6735
Error	1	0.008656	0.008656		
Total	2	0.011402			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.2408	5.29076	.0.093040	1.758533

β_i	df	b_i	$s(b_i)$	t*	Pr> t*
β_0	1	1.832633	0.142121		
β_1	1	-.037050	0.065789	-.5630	0.6634

β_i	CI_{lb}	b_i	CI_{ub}
β_0	0.050795	1.832633	3.614500
β_1	-.861880	-.037050	0.787780

F* = MSR/MSE F(.95,1,1) = 161 $t^* = b_1/s(b_1)$ t(.95,1) = 6.31

CI: $b_i \pm t(.95,1)s(b_i)$

Figure 3.27 AX: NEC's per individual - Regression results.

	77	78	79	80	81	82	83
AW				Y_{ij}			

MODEL: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

INDICES: i = year group

j = j^{th} individual from cell i ($j = 1, \ldots, 30$)

Y_{ij} = number of NEC's acquired by the j^{th} individual from cell i

μ = average number of NEC's per individual

τ_i = additional number of NEC's an individual from year group i receives

ε_{ij} = error terms that are iid $N(0,\sigma^2)$

We will follow the same outline presented earlier when we used the single factor ANOVA model to analyze the length of basic training. Our objectives for this section are:

- Estimate the mean number of NEC's per year group.
- Statistically test the means for differences.
- Rank the means using a paired comparison test.

The ANOVA model and the Kruskal-Wallis (KW) test will determine if the means differ. Tukey's paired comparison test will rank the means. The hypothesis tests associated with the Analysis of Variance model and the Kruskal-Wallis test are listed in Table XVII. The decision rules are also listed in Table XVII.

Test results, tables, and figures that support this subsection are grouped together. They are laid out in the following manner.

TABLE XVII

SINGLE FACTOR ANOVA HYPOTHESIS TEST #2

H_0: $\tau_{77} = \ldots = \tau_{83}$

H_1: $\tau_{77} \neq \ldots \neq \tau_{83}$

H_0: The mean number of NEC's per year group has remained constant.

H_1: Not all the means are equal.

-ANOVA-

If $F^* \leq F(.95, \nu_1, \nu_2)$ then conclude H_0

If $F^* > F(.95, \nu_1, \nu_2)$ then conclude H_1

-KW-

If $\chi^2_{KW} \leq \chi^2(.95, \nu)$ then conclude H_0

If $\chi^2_{KW} > \chi^2(.95, \nu)$ then conclude H_1

AT	Figure 3.28	Data Analysis Graphs
	Table XVIII	ANOVA/KW test results
	Figure 3.29	Tukey's paired comparison test results
AW	Figure 3.30	Data Analysis Graphs
	Table XIX	ANOVA/KW test results
	Figure 3.31	Tukey's paired comparison test results
AX	Figure 3.32	Data Analysis Graphs
	Table XX	ANOVA/KW test results
	Figure 3.33	Tukey's paired comparison test results

Figures 3.28, 3.30, and 3.32 provide a graphical summary of the data sets. Tables XVIII, XIX, and XX provide the ANOVA test results and the Kruskal-Wallis test results. Figures 3.29, 3.31, and 3.33 provide Tukey's paired comparison test results.

These figures display a graphical ranking of the means. Specific results are listed in the figures and tables. We summarize the findings.

- AT rating: ($F^* < F$) and ($\chi^2_{KW} < \chi^2$). By our decision rule, we accept H_0 and conclude, *"The mean number of NEC's acquired per year group has remained constant."*
- AW rating: The P value is .001. The test results are statistically significant. The elements of the τ vector are not equal. Using our decision rule, we accept the alternate hypothesis. Figure 3.31 provides a closer look at the differences. All means are grouped together under category A except year group 78. Those grouped together are not statistically different. Year Group 78 does not belong to group A, but look at the numbers. In particular, look at the largest mean (2.1), and look at the smallest mean (1.3). The difference is statistically significant but not operationally significant![2] We conclude by saying: *"A change occured but it is not operationally significant to influence training costs."*
- AX rating: ($F^* < F$) and ($\chi^2_{KW} < \chi^2$). By our decision rule, we accept H_0 and draw the same conclusion stated for the AT rating, *no increase.*

[2]We defined operationally significant as a factor of two or more. For first term enlistees, increasing the number of NEC's up to a factor two should have little effect on training costs. The Navy's C-schools should have the capacity to train more first terms enlistees. However, (2×1.3) = 2.6 which is fairly close to 2.1. There is a possibility that this change has more importance than we've given it.

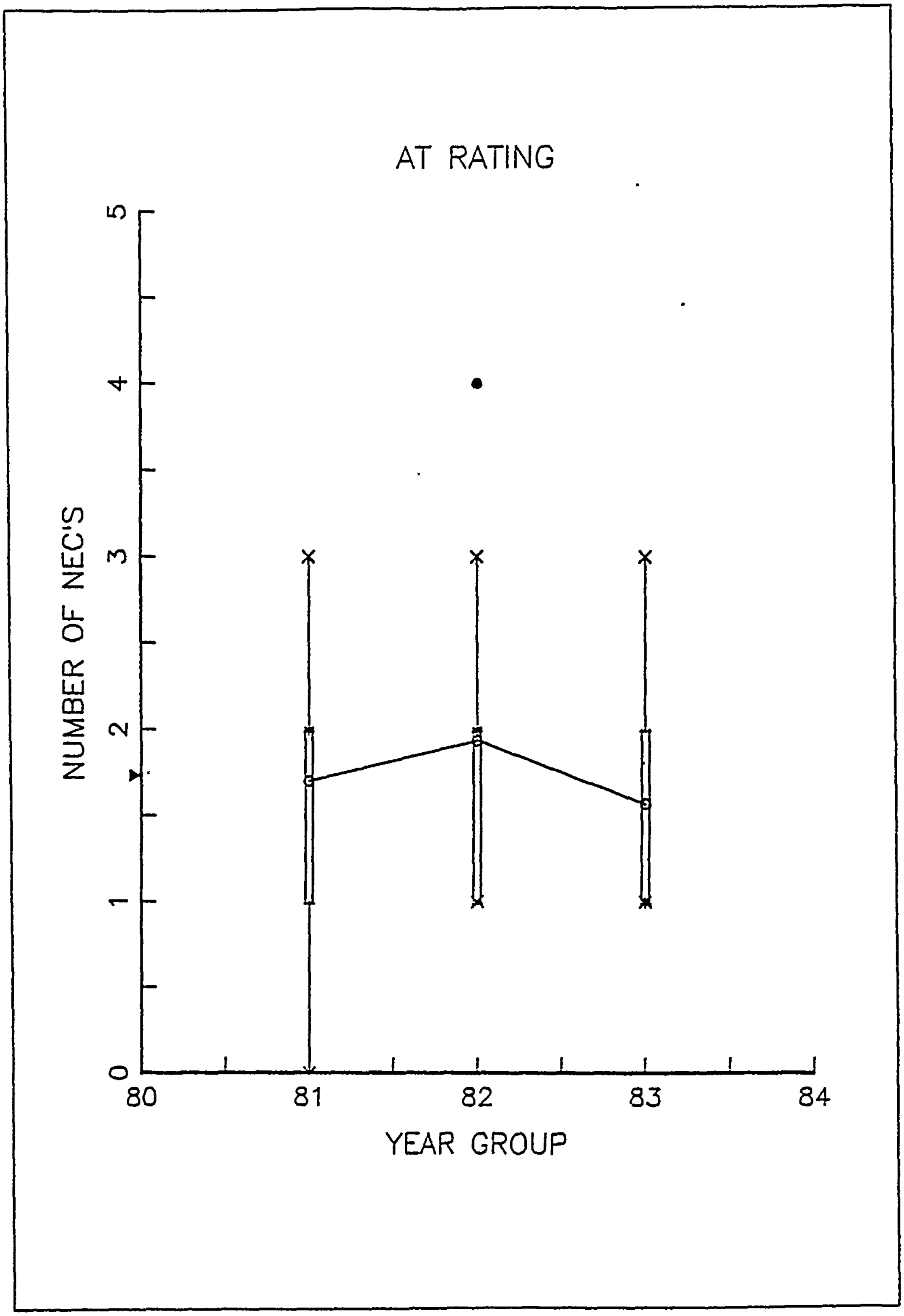


Figure 3.28 AT: NEC's per year group.

TABLE XVIII

AT: NEC'S PER YEAR GROUP
ANOVA TEST RESULTS

					-PERCENTILES-		
i	n_i	%	$\mu + \tau_i$	σ_i	0.25	0.50	0.75
81	30	33.3	1.700	0.6513	1	2	2
82	30	33.3	1.933	0.7397	1	2	2
83	30	33.3	1.567	0.6261	1	1	2
	90	100.0	1.733	0.6837	1	2	2

CLASS	LEVELS	VALUES
τ	3	81 82 83

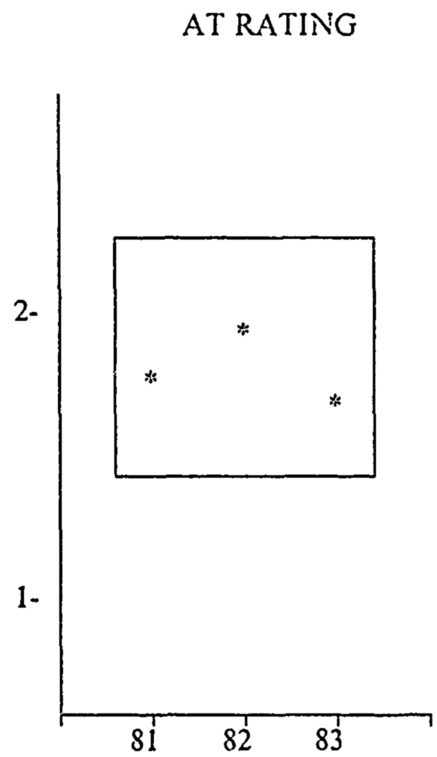
S	df	SS	MS	F*	PR> F*
Model	2	2.0667	1.0333	2.27	0.1089
Error	87	39.5333	0.4544		
Total	89	41.6000			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0497	38.8902	0.6740	1.7333

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 2)$
2	4.1309	0.1268

$F(.95,2,87) = 3.11 \quad \chi^2(.95,2) = 5.99$



i	n_i	$\mu + \tau_i$
81	30	1.7000
82	30	1.9333
83	30	1.5667

Means boxed together are not statistically different.

Figure 3.29 AT: Tukey's paired comparison test results #2.

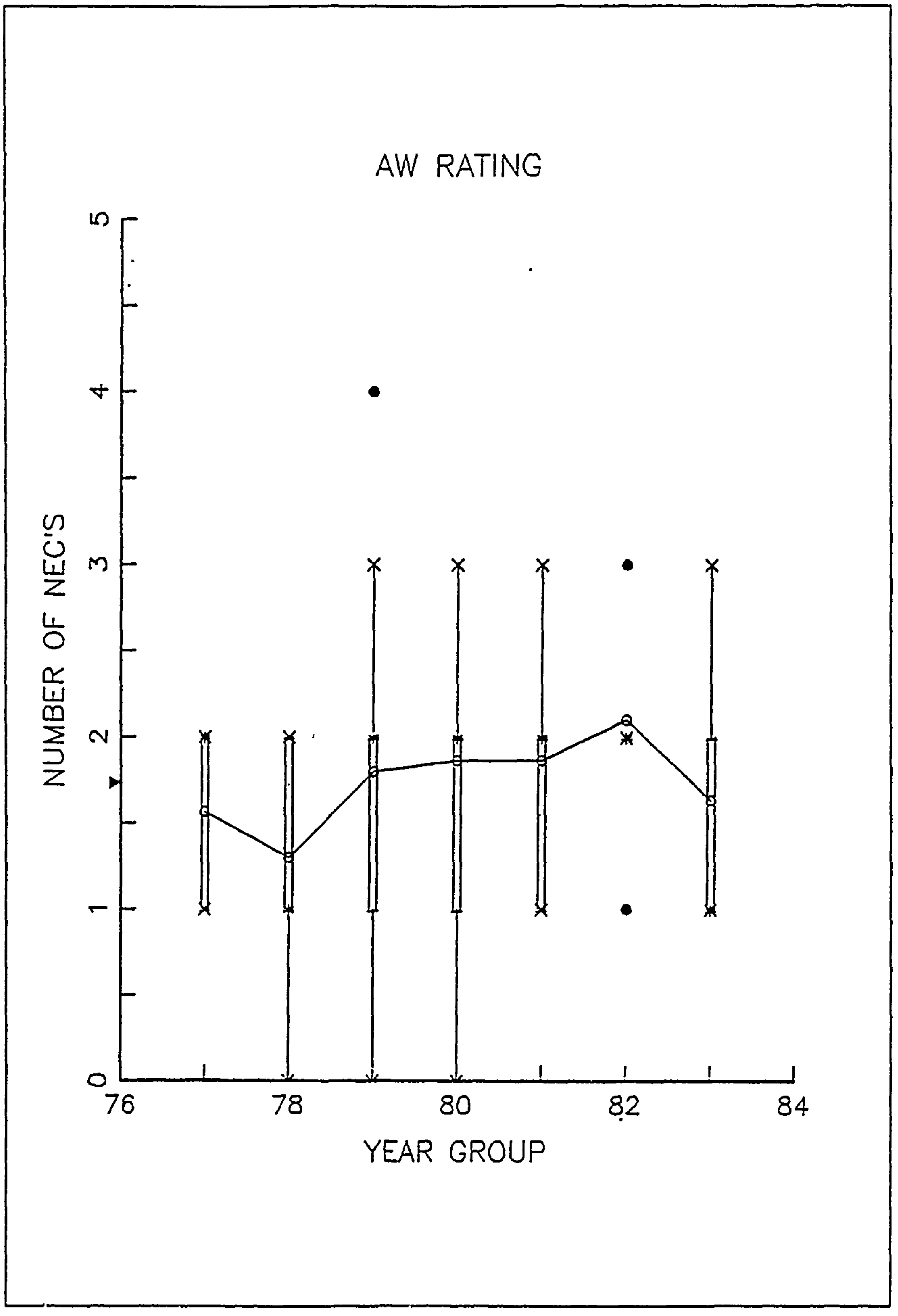


Figure 3.30 AW: NEC's per year group.

TABLE XIX

AW: NEC'S PER YEAR GROUP
ANOVA TEST RESULTS

i	n_i	%	$\mu + \tau_i$	σ_i	-PERCENTILES- 0.25	0.50	0.75
77	30	14.3	1.567	0.5940	1	2	2
78	30	14.3	1.300	0.6513	1	1	2
79	30	14.3	1.800	0.9966	1	2	2
80	30	14.3	1.867	0.8604	1	2	2
81	30	14.3	1.867	0.6815	1	2	2
82	30	14.3	2.100	0.6074	2	2	2
83	30	14.3	1.633	0.7184	1	1	2
	210	100.0	1.733	0.7611	1	2	2

CLASS	LEVELS	VALUES
τ	7	77 78 79 80 81 82 83

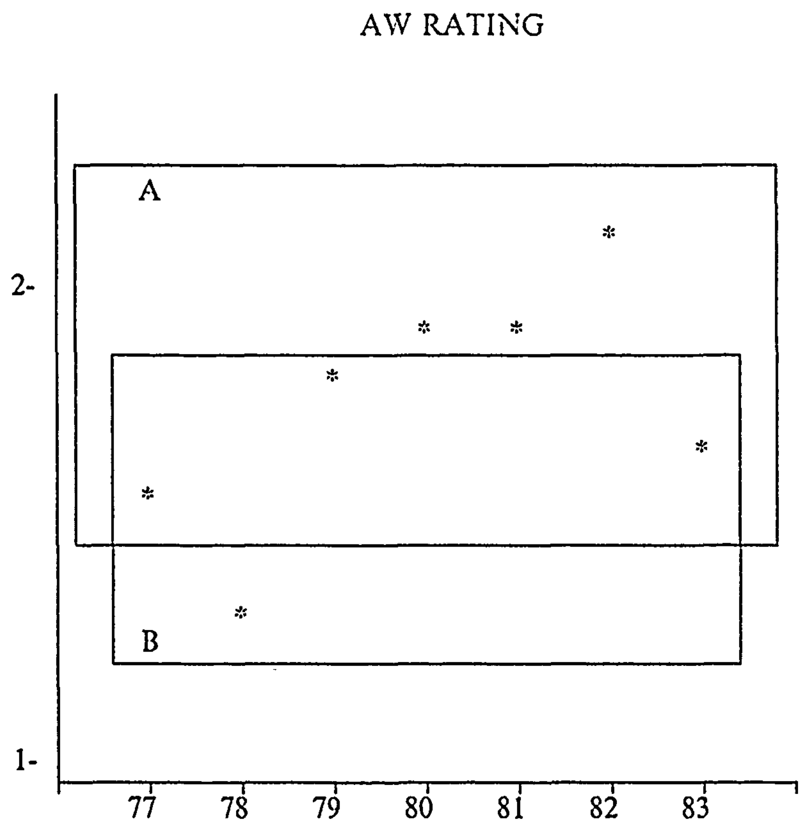
S	df	SS	MS	F*	PR> F*
Model	6	12.0000	2.0000	3.73	0.0016
Error	203	109.0667	0.5372		
Total	209	121.0667			

R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0991	42.2879	0.7330	1.7333

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 6)$
6	21.65	0.0014

$F(.95,6,203) = 2.10 \quad \chi^2(.95,6) = 12.59$



i	n_i	$\mu + \tau_i$
77	30	1.5667
78	30	1.3000
79	30	1.8000
80	30	1.8667
81	30	1.8667
82	30	2.1000
83	30	1.6333

Means boxed together are not statistically different.

Figure 3.31 AW: Tukey's paired comparison test results #2.

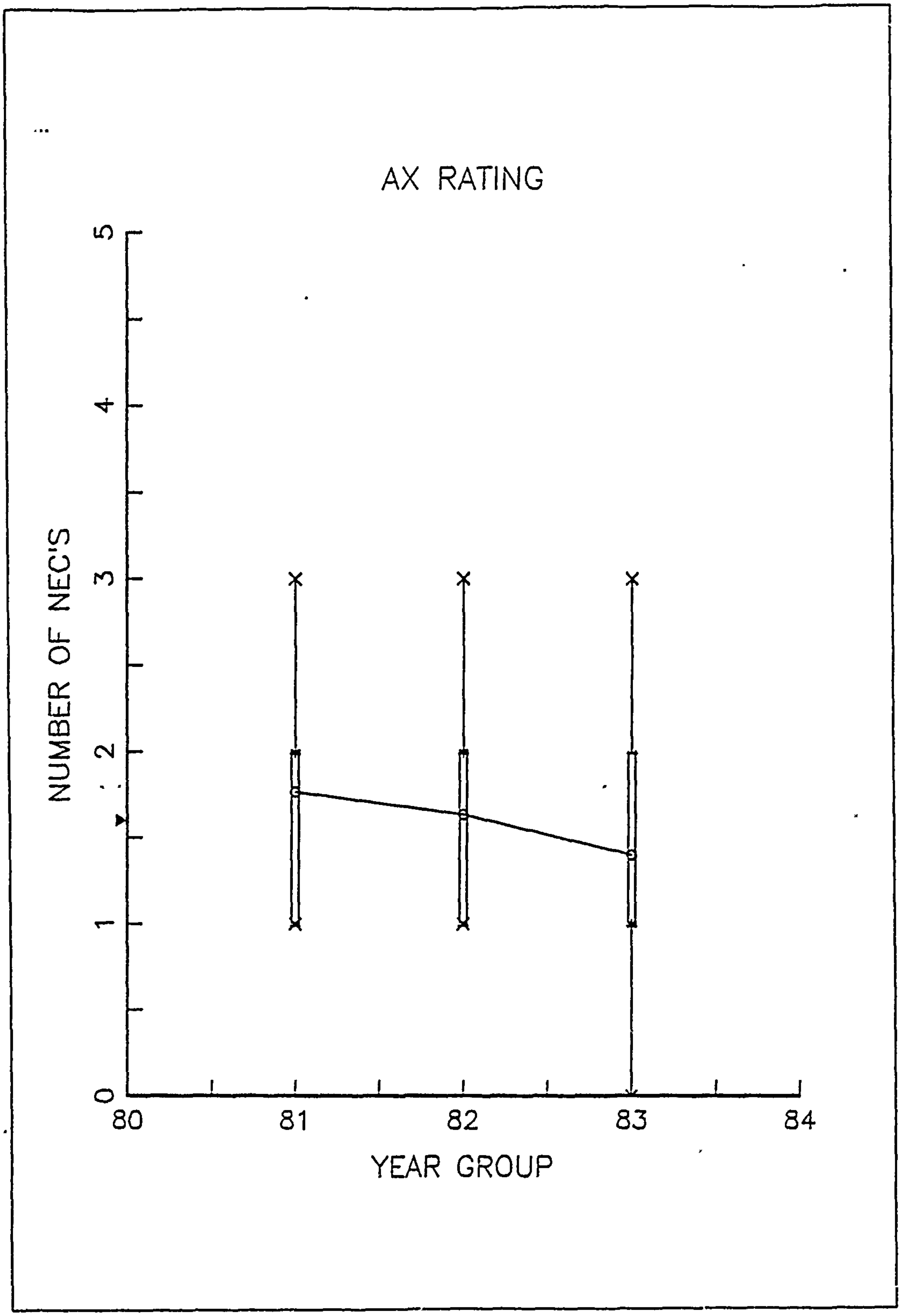


Figure 3.32 AX: NEC's per year group.

TABLE XX

AX: NEC'S PER YEAR GROUP
ANOVA TEST RESULTS

i	n_i	%	$\mu + \tau_i$	σ_i	-PERCENTILES- 0.25	0.50	0.75
81	30	33.3	1.767	0.7279	1	2	2
82	30	33.3	1.833	0.5560	1	2	2
83	30	33.3	1.400	0.6215	1	1	2
	90	100.0	1.600	0.6500	1	2	2

CLASS	LEVELS	VALUES
τ	3	81 82 83

S	df	SS	MS	F*	PR > F*
Model	2	2.0667	1.0333	2.53	0.0855
Error	87	35.5333	0.4084		
Total	89	37.6000			

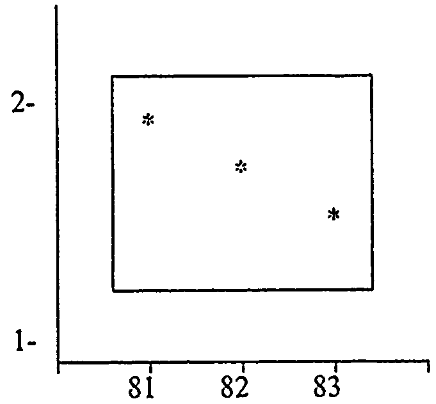
R^2	C.V.	$\sqrt{MSE}$	μ_Y
0.0550	39.9428	0.6391	1.6000

KRUSKAL-WALLIS NONPARAMETRIC TEST FOR EQUAL MEANS

df	χ^2_{KW}	PR > $\chi^2(.95, 2)$
2	4.26	0.1186

$F(.95,2,1109) = 3.00 \quad \chi^2(.95,2) = 5.99$

AX RATING



i	n_i	$\mu + \tau_i$
81	30	1.7667
82	30	1.6333
83	30	1.4000

Means boxed together are not statistically different.

Figure 3.33 AX: Tukey's paired.comparison test results #2.

IV. MAIN RESULTS AND CONCLUSIONS

We started off with the following question, "What are the factors causing training costs to rise?" To understand the problem, we formulated several reasons why we think training costs are rising. Those reasons are:

- The length of basic training has increased.
- Attrition has increased.
- The amount of specialized training has increased.

We set out to verify those reasons using some historical data compiled by CNA.

The scope of this study is limited. The results are valid within the following confines.

- Inferences are made with respect to these enlisted ratings, AT, AW, and AX.
- The expected career path is Boot Camp → A-School → Fleet. Inferences are further restricted to those individuals that followed the expected career path.
- The overall time frame is restricted to the first enlistment period.
- The first 24 months is the time constraint for two areas of study, Basic Training and Attrition.
- The second and third years of service is the time constraint for the last area of study, Specialized Training.

A. SUMMARY

1. (*Length of Basic Training — not a factor*) The length of basic training has cycled up and down. It has fluctuated over the years but there is no evidence to suggest a steady increase over the years. Figures 4.1, 4.2, and 4.3 provide graphical summaries. In all three cases, the final trend is encouraging, the length of basic training has decreased.
2. (*Attrition — not a factor*) Losses in Basic Training are roughly constant from year to year. Attrition has not increased.
3. (*Amount of specialized training — not a factor*) Specialized training has remained constant. The amount has not increased.

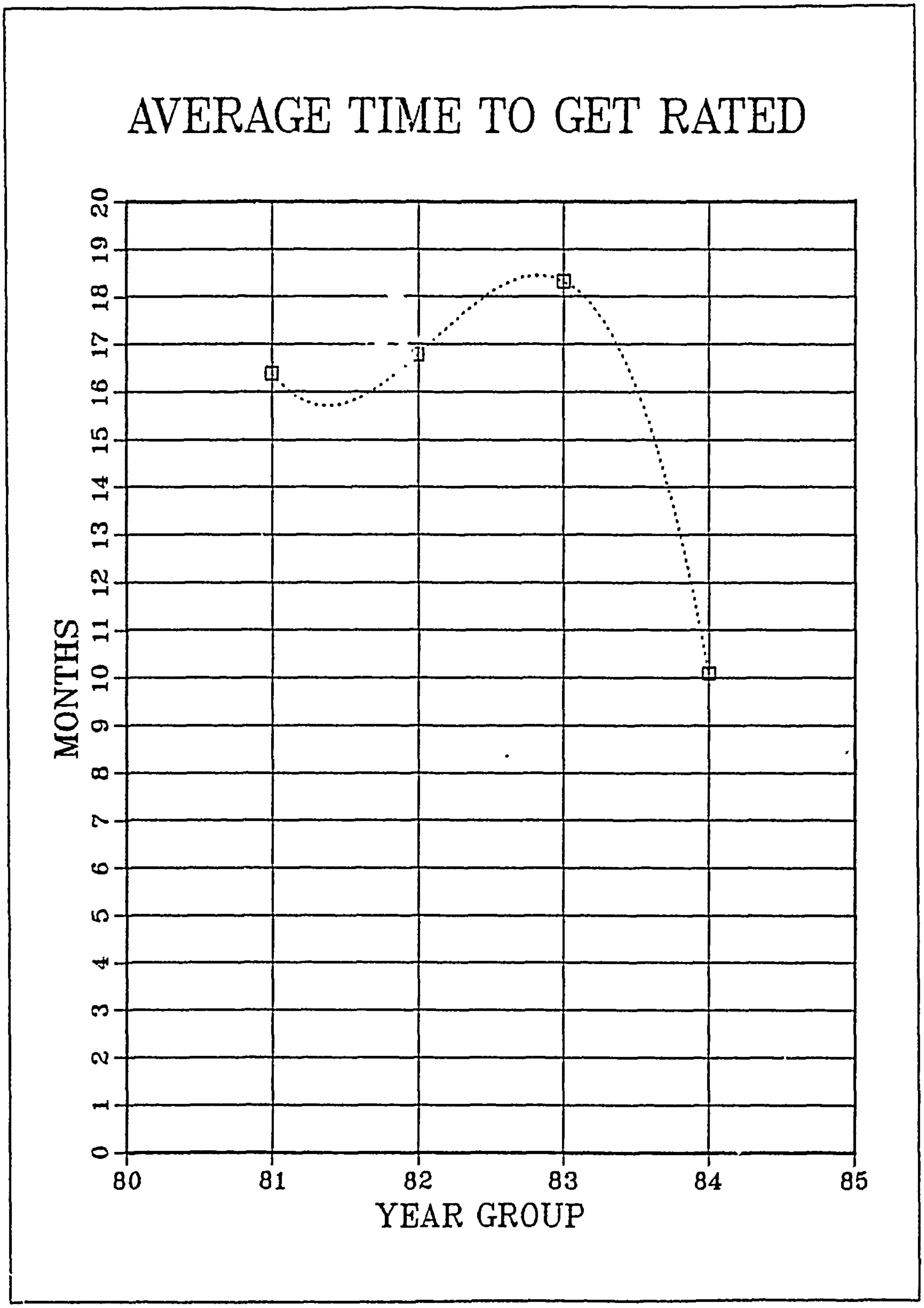


Figure 4.1 AT: Length of basic training.

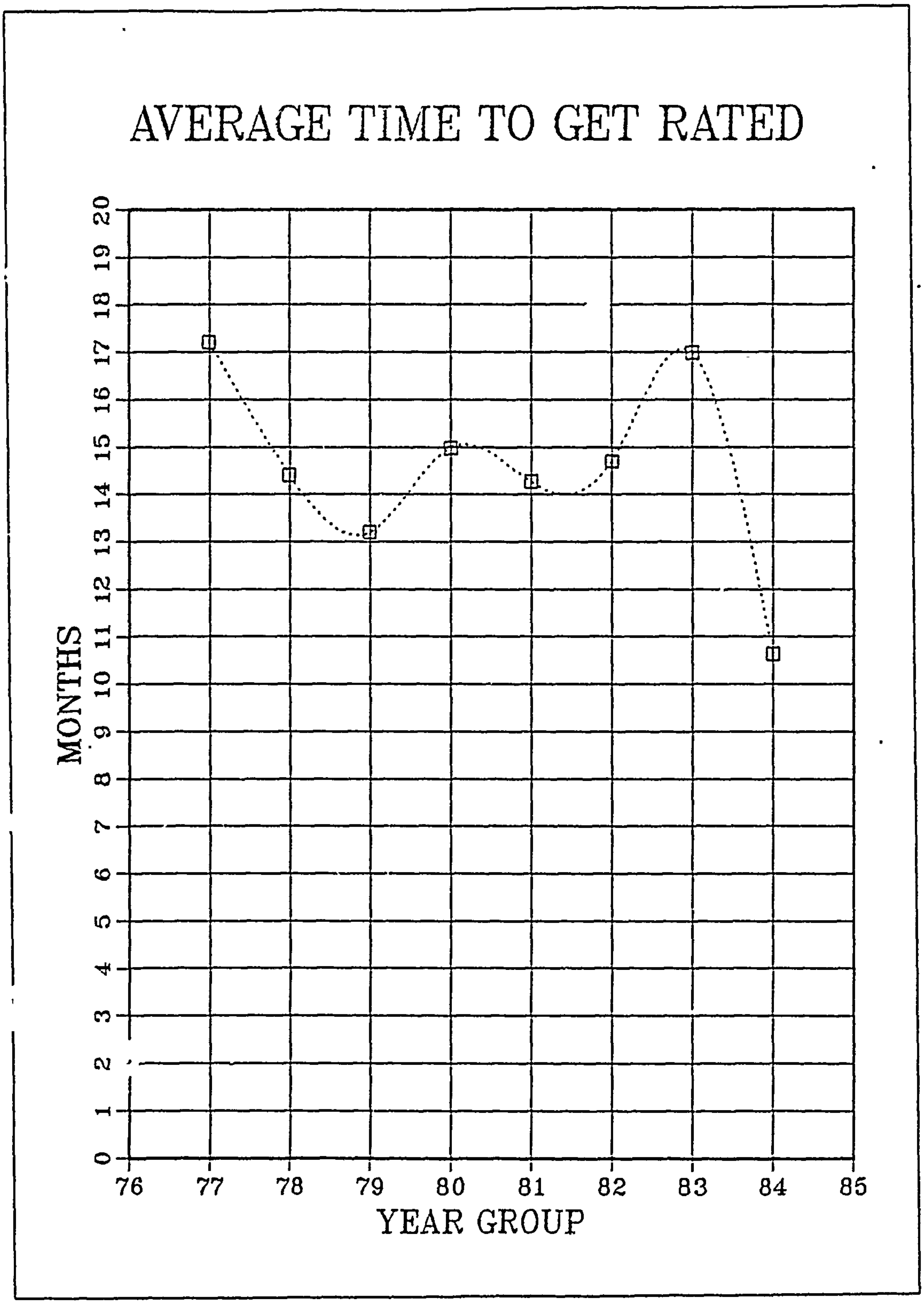


Figure 4.2 AW: Length of basic training.

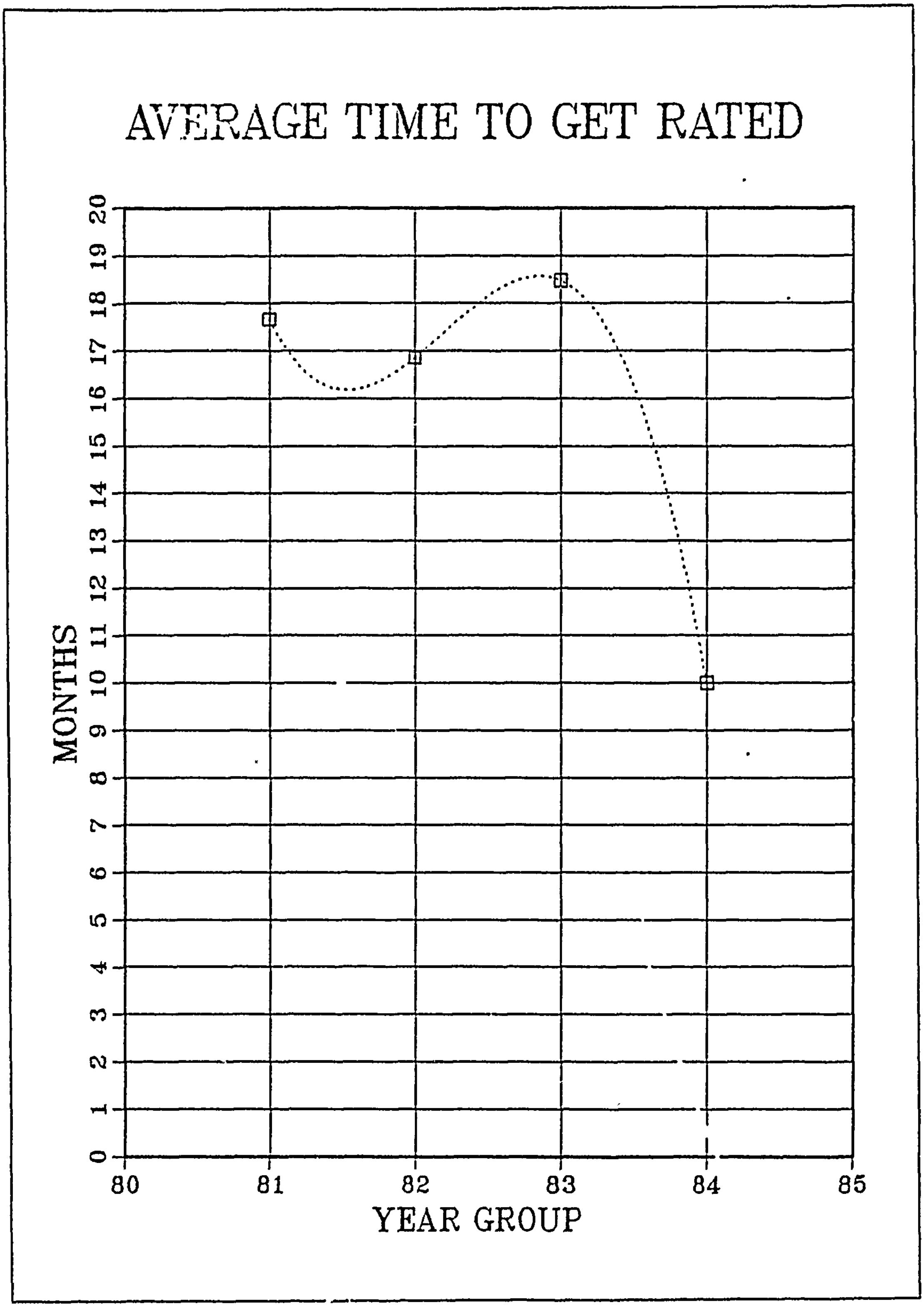


Figure 4.3 AX: Length of basic training.

B. RECOMMENDATIONS

This study looked at a small piece of the problem. The final result is that we were unable to identify any factors causing training costs to rise. However, here is a list of general questions that may be of interest for further research.

1. Has the length of basic training increased for enlisted ratings other than AT, AW, and AX?
2. Has the amount of specialized training increased after the first enlistment period?
3. Is the selection process effective?[3]
4. Has the Training Command's support costs increased?
5. Are training costs rising due to increased or improved training resources?

This list is by no means exhaustive. It is a few questions that we can ask but were unable to answer in t. i. .'''dy.

[3]The selection process is primarily based upon test scores and education level. If the selection process is effective, then people screened for a particular rating will complete that training program. The attrition rate will be low and survivability high. However, if we do not screen people properly, the number of people that complete the program will be much less than optimal. Attrition will be high. The effect is higher training costs. An effective selection process produces savings.

APPENDIX A
MODEL ASSUMPTIONS

Throughout this study, we used two models extensively, the REGRESSION model and the ANOVA model. Both models helped us to conceptualize the problem and analyze the observations. The purpose of both models is to describe the events of the past. These models are also used to predict and control events, but we're not interested in using it for those matters.

In this appendix, we will briefly assess the aptness of the model. Is the model appropriate for the data set at hand? This is an important question. It should be answered whenever models are used. The importance of aptness is best described by *logic's implication statement*, if P then Q, (P $\rightarrow$ Q). If the model is appropriate, then the ensuing analysis presented by the model is correct. Good analysis is conditioned on the fact that the analyst use the appropriate models. The appropriateness of a model is dependent upon adherence to the assumptions imbedded within the model.

We emphasized the importance of examining the aptness of a model, but how do we confirm that a model is appropriate? Residual analysis is the tool for this task. It is highly effective for spotting major departures from the assumed model. Our goal is to verify the model assumptions by using residual analysis. In the statistical world, this verification follows the mentality used in the U.S. court system, where we assume the defendant to be innocent and prove beyond reason of doubt that the person is guilty. In our profession, we assume the model assumptions are correct and prove otherwise. The major purpose of residual analysis is to detect serious departures from the conditions assumed by the model.

Strict adherence to every assumption is not possible with this data set. A few departures exist however, the departures are not substantial. Our first discussion centers around the regression model. The second part deals with the ANOVA model. Assumptions are listed for each model. This is followed by a short summary discussing the verification procedures and any effects caused by a departure from the model. Figures and tables pertain to the AW rating. Similar results were obtained for the AT and AX ratings.

1. REGRESSION

We used graphical means to confirm the assumptions imbedded within the regression model. (See Table XXI.) The assumptions are listed in column one. The plots used to confirm these assumptions are listed in column two. Our goal is to ensure the assumptions are plausible in light of the data.

TABLE XXI
REGRESSION MODEL ASSUMPTIONS

Assumption	Verification
1. The relationship is linear.	Scatter Plot
2. The error terms are independent.	RESID vs X RESID vs YHAT
3. The error terms have constant variance.	RESID vs X RESID vs YHAT
4. The error terms are normally distributed.	Q-Q Plot

a. The relationship is linear.

Whether or not a linear regression function is appropriate for the data set at hand being analyzed, can often be studied by a scatter plot of the data. (See Figure A.1.) These scatter plots are an effective means to examine the appropriateness of the linear regression function. Notice that these plots do not exhibit any departures from the model.

b. The errors are independent and have constant variance.

If the model correctly describes the observations, the (RESID vs X) plot and the (RESID vs YHAT) plot should display a pattern that's uniformly distributed within a horizontal band centered at zero. (See Figures A.2 and A.3.) It portrays the prescribed behavior. No trends are present.

c. **The error terms are normal.**

The residuals should resemble observations taken from a normal distribution. The Q-Q plots are used to confirm this. Figure A.4 displays these plots. They appear to be normally distributed.

In summary, no serious departures from the assumptions were noted. *The linear regression model is appropriate for the data set at hand.*

2. ANALYSIS OF VARIANCE

The assumptions imbedded within the ANOVA model are similar to the regression model. See Table XXII for a list of the assumptions and the verification method.

TABLE XXII

ANOVA MODEL ASSUMPTIONS

Assumption	Verification
1. The populations are normally distributed.	
2. The population variances are equal.	Bartlett Test Hartley Test
3. The error terms are independent.	Durbin-Watson Test
4. The error terms have constant variance.	RESID vs X RESID vs YHAT
5. The error terms are normally distributed.	Histogram

a. **The populations are normally distributed.**

The first assumption requires the populations to be normally distributed. Formal verification will not be presented here. It will suffice to say that upon examination of the data sets, we found most of the populations to lack normality. Here in lies the first departure from the model, but the departure is not large. Lack of normality is not an important matter provided the departure from normality is not of extreme form. The point estimators of factor level means and contrasts are unbiased whether or not the populations are normal. The F-test for equality of means is but little affected by lack of normality, either in terms of level of significance or power of the test. Hence the F-test is a *robust* test against departures from normality. [Ref. 2: p. 624]

b. **The population variances are equal.**

The second assumption requires equal variances. We used the Bartlett test or the Hartley test to verify homogeniety of variance. Let's discuss where we applied each test.

1. Basic Training: Bartlett Test - (unequal sample sizes)

The idea underlying Bartlett's test[4] is simple. By definition:

$$MSE = (1/df_T)\sum df_i\, s_i^2 \qquad \text{(eqn A.1)}$$

$$GMSE = [\, (s_1^2)^{df_1} \times \ldots \times (s_n^2)^{df_n} \,]^{(1/df_T)} \qquad \text{(eqn A.2)}$$

The relationship between the arithmetic mean and the geometric mean is:

$$GMSE \leq MSE \qquad \text{(eqn A.3)}$$

The two averages will be equal if $s_i = s_j$, hence if the ratio (MSE/GMSE) is close to one, we have evidence the variances are equal. If the ratio is large, it indicates that the population variances are unequal. Bartlett's test statistic is computed as follows:

$$\chi^2_B = (\, df_T/C \,)(\, \log_e MSE - \log_e GMSE \,) \qquad \text{(eqn A.4)}$$

where:

$$C = 1 + [1/3(n - 1)]\, \{\, [\, \sum(1/df_i)\,] - (1/df_T)\, \} \qquad \text{(eqn A.5)}$$

[4][Ref. 2: Sec. 18.6] provides a detailed discussion of this test.

The population variances are listed in Table XXIII. We statistically tested these values to degermine if they were equal. The hypothesis test and decision rule associated with Bartlett's test are listed in Table XXIII. The results are also listed in Table XXIII.

With respect to the AW and AX ratings, we accept the null hypothesis and conclude, the population variances are equal. However, we cannot say the same for the AT rating. Departure from this model assumption has some effect. How sensitive is the model with respect to this departure?

When the error variances are unequal, the F-test for equality of means is only slightly affected if all factor level sample sizes are equal or do not differ greatly. Specifically, unequal error variances raise the actual level of significance only slightly higher than the specified level. The F-test is robust against unequal variances when the sample sizes are approximately equal. [Ref. 2: p. 624].

Let's look at this aspect more closely. For the AT rating, the population variances are unequal and the sample sizes are unequal. We expect the significance level to be inflated. However, if a large inflation factor existed, it would not have affected this ANOVA test very much. This is due to the fact that the test results were significant at the .0001 level! The difference in means is causing the significance level to be extremely small. It's overpowering any inflationary effect caused by unequal variances. The actual probability that the means are equal is somewhat less than .0001. In summary, a departure from the model is present, the population variances are not equal. However, this does not bias the true results very much. In this case, we accept the validity of the F-test results.

2. *Specialized Training*: *Hartley Test - (equal sample sizes)*

For equal sample sizes, Hartley's test[5] for equality of variance is based solely on the largest sample variance and the smallest sample variance. Hartley's test statistic is defined as follows:

$$H^* = \max(s_i^2)/\min(s_j^2) \qquad \text{(eqn A.6)}$$

[5][Ref. 2: Sec. 18.6] provides a detailed discussion of this test.

Clearly, values of H* near one support the claim that the population variances are equal. The variances for each population are listed in Table XXIV. The hypothesis test and decision rule associated with the Hartley test are listed in Table XXIV. The results are also listed in Table XXIV. For all three test cases, we conclude the population variances are equal.

c. **The error terms are independent.**

The third assumption requires the error terms to be independent. Lack of independence can have serious effects on the inferences made using the ANOVA output. The observations were obtained in time sequence, so there is a good chance the error terms are serially correlated or autocorrelated.

The most popular test for first-order autoregressive errors is the Durbin-Watson (D-W) test. It's a powerful test yet extremely easy to use. See [Ref. 5: Sec. 15.3] for a detailed commentary on the (D-W) statistic. The original model specifies the error terms (ε_i) to be independent and identically distributed $N(0,\sigma^2)$ random variables. The underlying arguement for the D-W test is simple. Model the error term as a first-order autoregressive process such that:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \upsilon_i \qquad \text{(eqn A.7)}$$

where:

ρ = autocorrelation parameter such that $|\rho| < 1$

υ_i = disturbance terms that are iid $N(0,\sigma^2)$

Each error term includes a fraction of the previous error term plus a disturbance term. If $\rho = 0$, then $\varepsilon_i = \upsilon_i$, and we're back to our original assumption because the disturbance terms (υ_i) are independent. The D-W test determines if $\rho = 0$. The hypothesis test and decision rule associated with the D-W test are listed in Table XXV. The Durbin-Watson test results are also listed in Table XXV. For every test case, we conclude: *"The autocorrelation parameter ρ is zero hence, the error terms are independent."*

d. **The error terms have constant variance.**

Assumption number four requires the error terms to have constant variance. (See Figures A.5 through A.8.) We ploted the residuals against the independent variable and the fitted value. No discernable pattern emerged. The residuals lie within a horizontal band centered at zero. Notice how the variance stays constant through changes on the X-axis. This behavior is the expected behavior given the assumption is correct. These plots give us no reason to reject the fourth assumption.

e. **The error terms are normally distributed.**

The last assumption requires the error terms to be normally distributed. We plotted the residuals in the form of a histogram. (See Figures A.9 and A.10.) Both plots resemble a normal distribution with mean zero. These plots verify the last assumption.

In summary, the assumptions are reasonable. We have no reason to reject them as incorrect. There is a few minor departures from the model, but due to the robustness of the F-test, these departures did not affect the final results. We conclude: *The ANOVA model is appropriate for the data set at hand.*

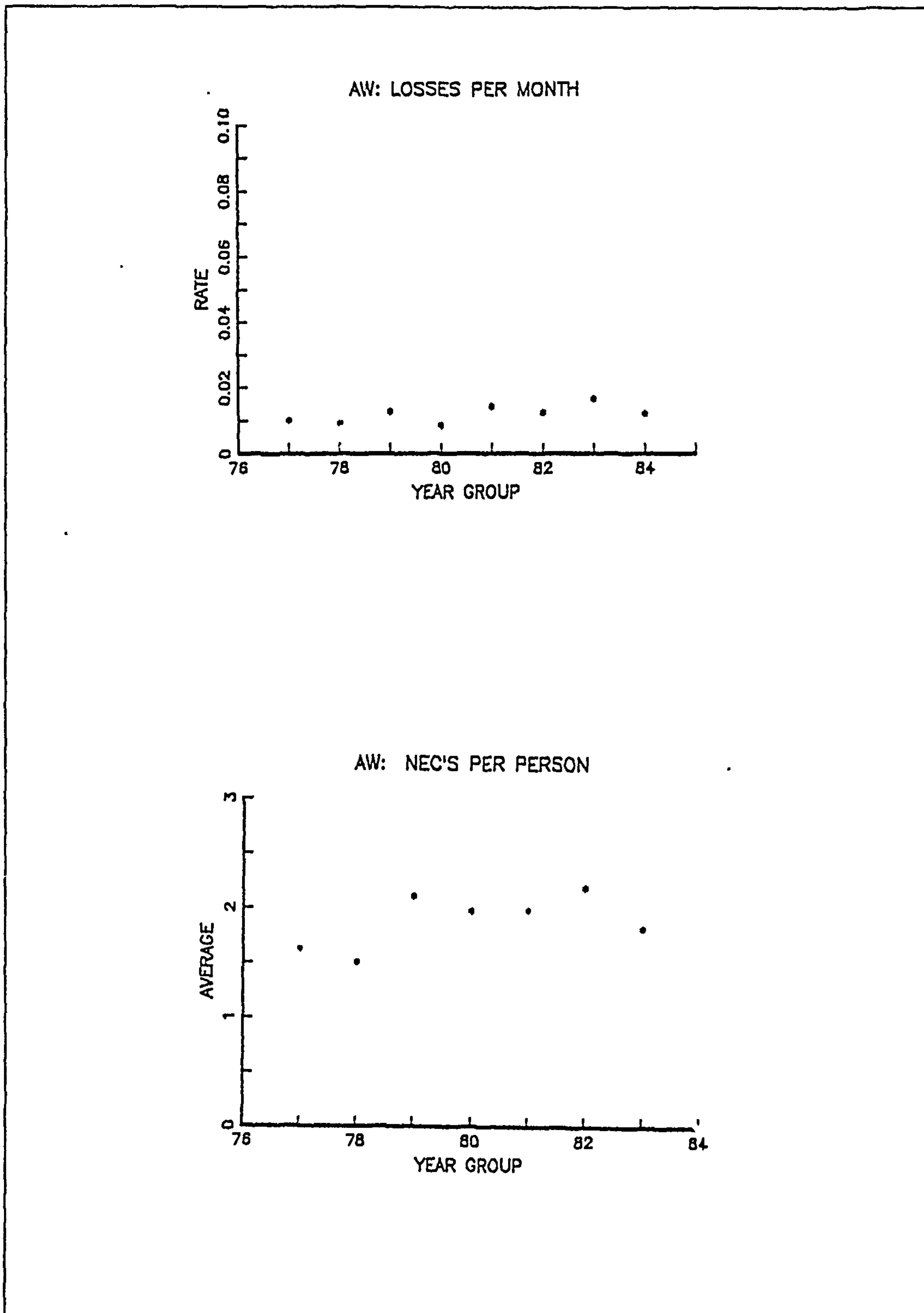


Figure A.1 AW Regression: Scatter Plot.

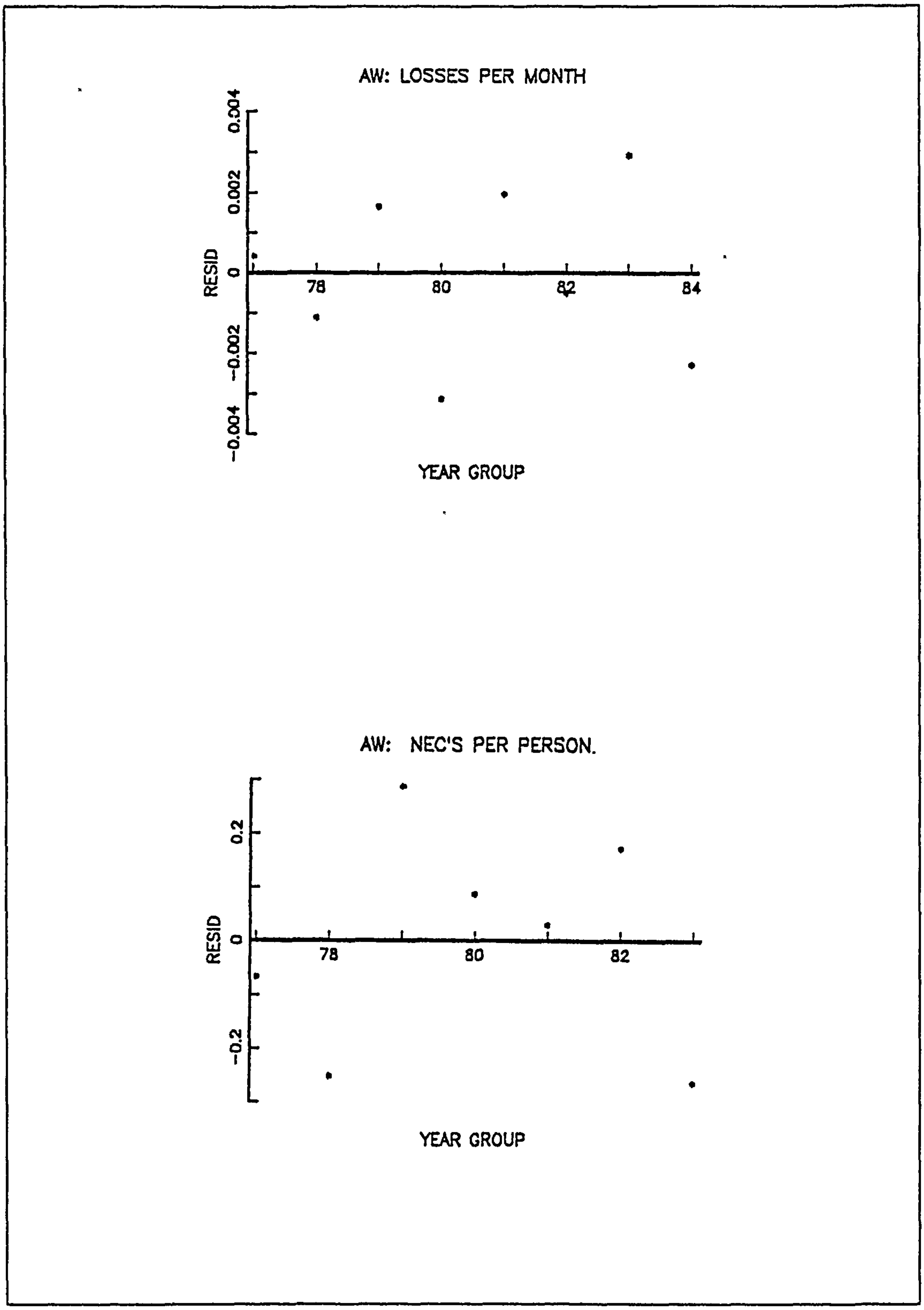


Figure A.2 AW Regression: RESID vs X Plot.

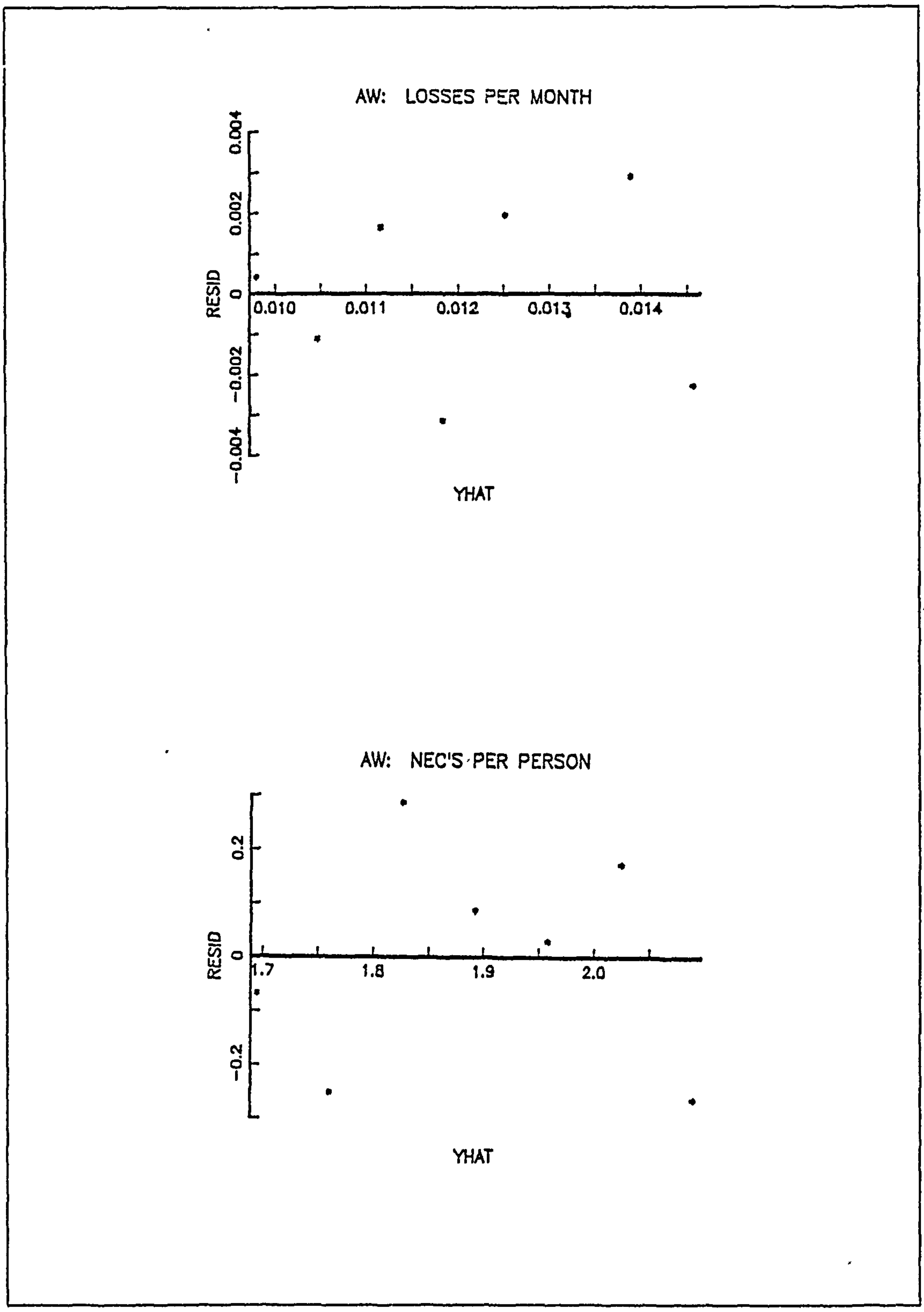
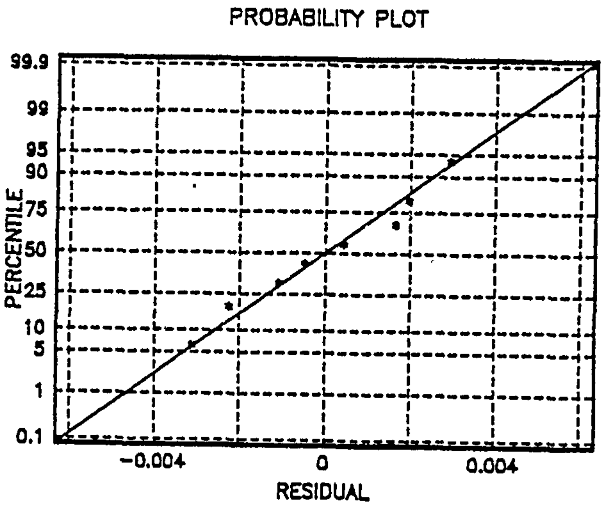


Figure A.3 AW Regression: RESID vs YHAT Plot.



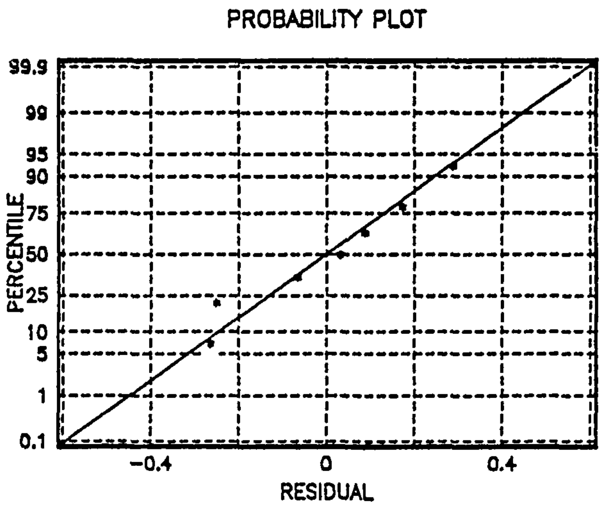


Figure A.4 AW Regression: Q-Q Plot.

TABLE XXIII

TIME TO GET RATED
BARTLETT'S TEST

	i	μ_i	df_i	σ_i
AT	81	16.385	225	3.5600
	82	16.785	520	2.7153
	83	16.385	364	3.1104
AW	77	17.214	69	3.1249
	78	14.414	98	3.7743
	79	13.199	160	3.7895
	80	14.986	208	3.9559
	81	14.270	173	3.8829
	82	16.703	302	3.2781
	83	16.988	242	3.2721
AX	81	17.667	32	3.4157
	82	16.863	138	2.9073
	83	18.481	78	2.7496

H_0: $\sigma_1^2 = \ldots = \sigma_n^2$
H_1: $\sigma_1^2 \neq \ldots \neq \sigma_n^2$

If $\chi^2_B \leq \chi^2(.95, \nu)$ then conclude H_0
If $\chi^2_B > \chi^2(.95, \nu)$ then conclude H_1

	ν	χ^2_B	$\chi^2(.95, \nu)$	$\therefore$
AT	2	6.2084	5.9915	H_1
AW	6	4.7408	12.5916	H_0
AX	2	0.0598	5.9915	H_0

TABLE XXIV

NEC'S PER YEAR GROUP
HARTLEY'S TEST

	i	μ_i	df_i	σ_i
AT	81	1.700	30	0.6513
	82	1.933	30	0.7397
	83	1.567	30	0.6261
AW	77	1.567	30	0.5940
	78	1.300	30	0.6513
	79	1.800	30	0.9966
	80	1.867	30	0.8604
	81	1.867	30	0.6815
	82	2.100	30	0.6074
	83	1.633	30	0.7184
AX	81	1.767	30	0.7279
	82	1.833	30	0.5560
	83	1.400	30	0.6215

H_0: $\sigma_1^2 = \ldots = \sigma_n^2$

H_1: $\sigma_1^2 \neq \ldots \neq \sigma_n^2$

If $H^* \leq H(.95, \nu_1, \nu_2)$ then conclude H_0

If $H^* > H(.95, \nu_1, \nu_2)$ then conclude H_1

	ν_1	ν_2	H^*	$H(.95, \nu_1, \nu_2)$	$\therefore$
AT	3	29	1.3958	2.4000	H_0
AW	7	29	2.8149	3.0200	H_0
AX	3	29	1.7139	2.4000	H_0

TABLE XXV

DURBIN-WATSON TEST

H_0: $\rho = 0$

H_1: $\rho > 0$

If DW > d_{ub} then conclude H_0

If DW < d_{lb} then conclude H_1

If $d_{lb} \leq DW \leq d_{ub}$ then the test is inconclusive

Time to get rated

Single Factor ANOVA Model

	DW*	d_{lb}	d_{ub}	$\therefore$
AT	2.022	1.758	1.778	H_0
AW	2.004	1.758	1.778	H_0
AX	2.080	1.758	1.778	H_0

NEC's per year group

Single Factor ANOVA Model

	DW*	d_{lb}	d_{ub}	$\therefore$
AT	2.210	1.635	1.679	H_0
AW	2.040	1.758	1.778	H_0
AX	2.180	1.635	1.679	H_0

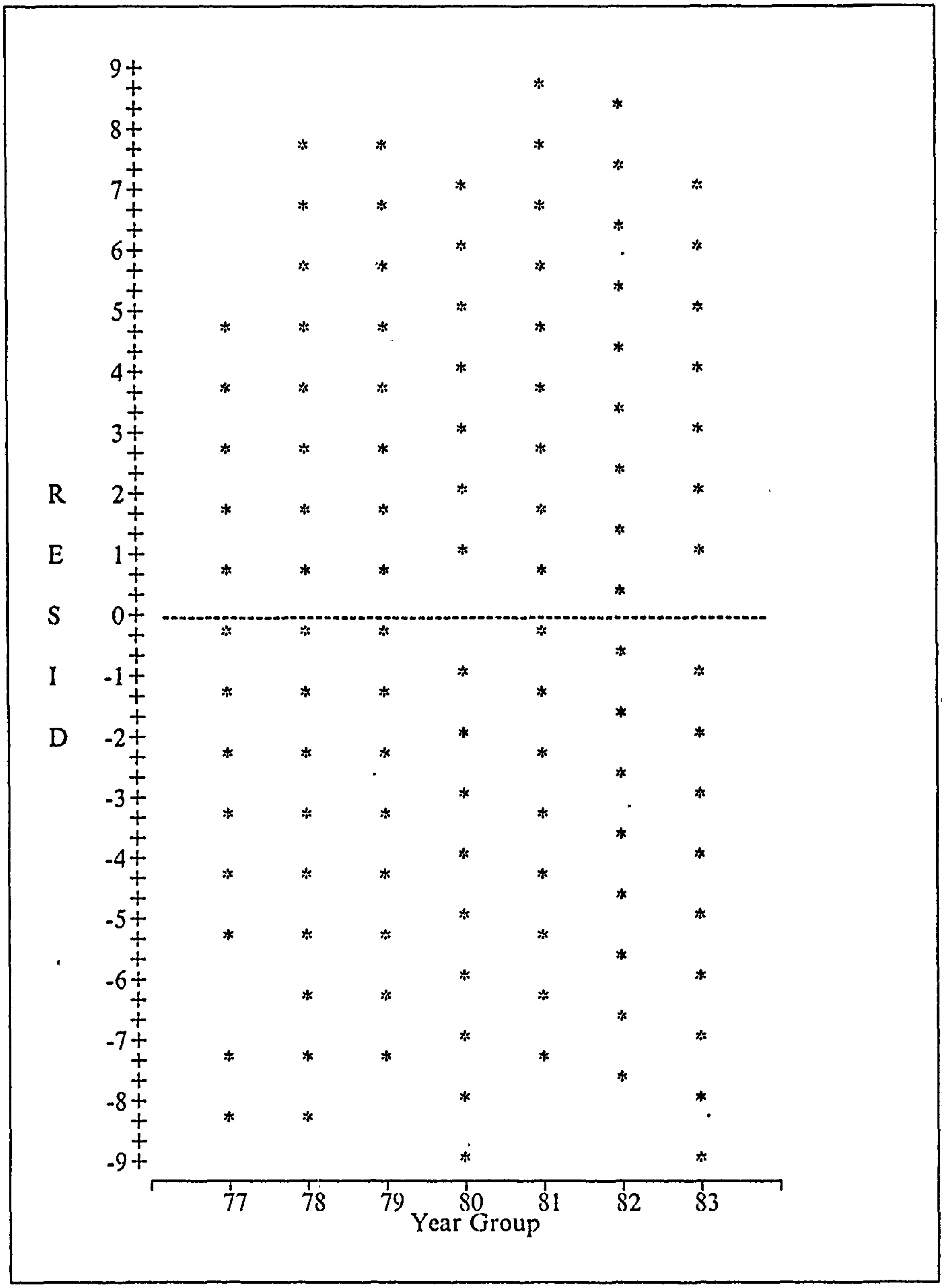


Figure A.5 AW ANOVA: Time to get rated - RESID vs X.

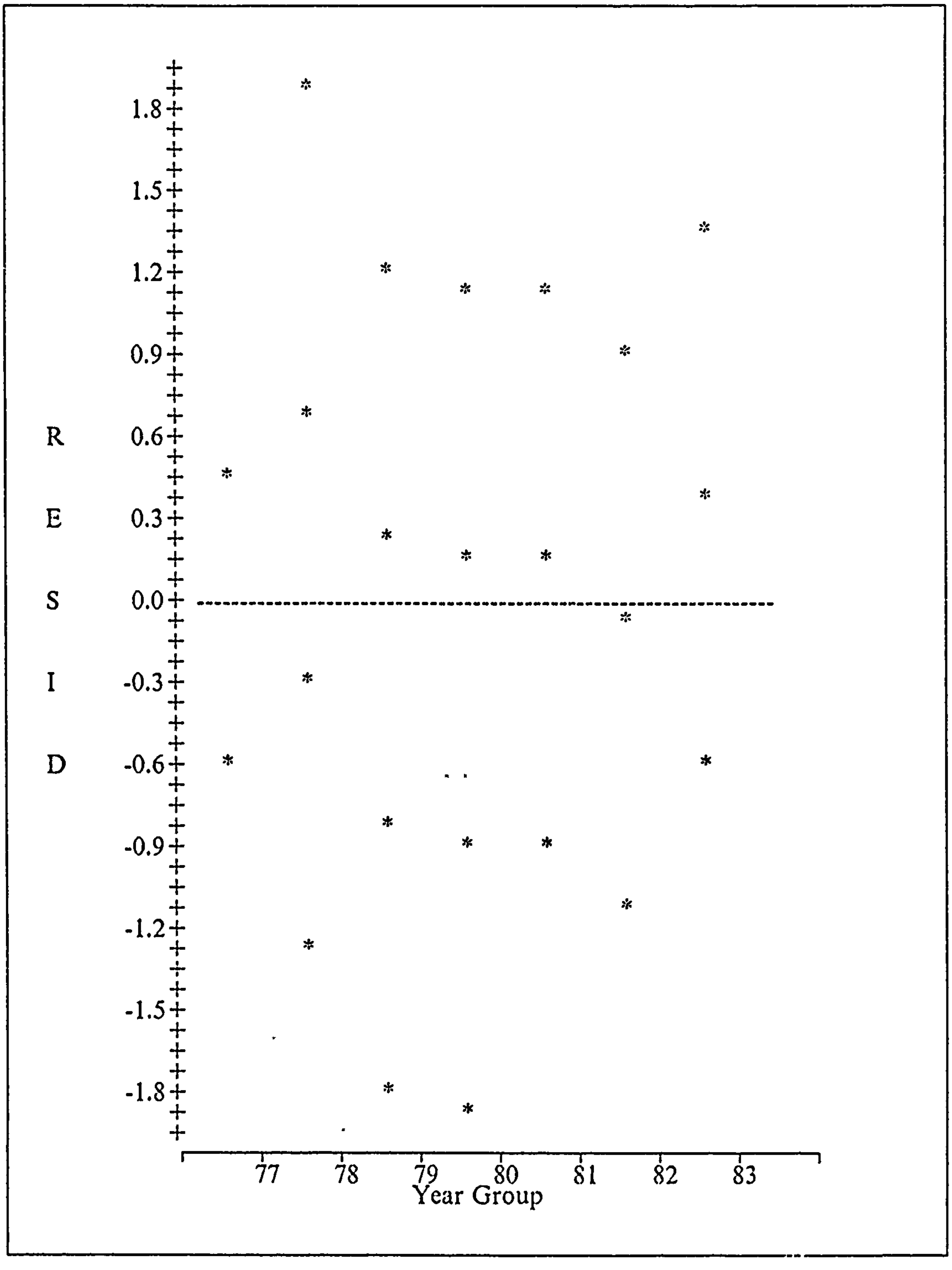


Figure A.6 AW ANOVA: NEC's per year group - RESID vs X.

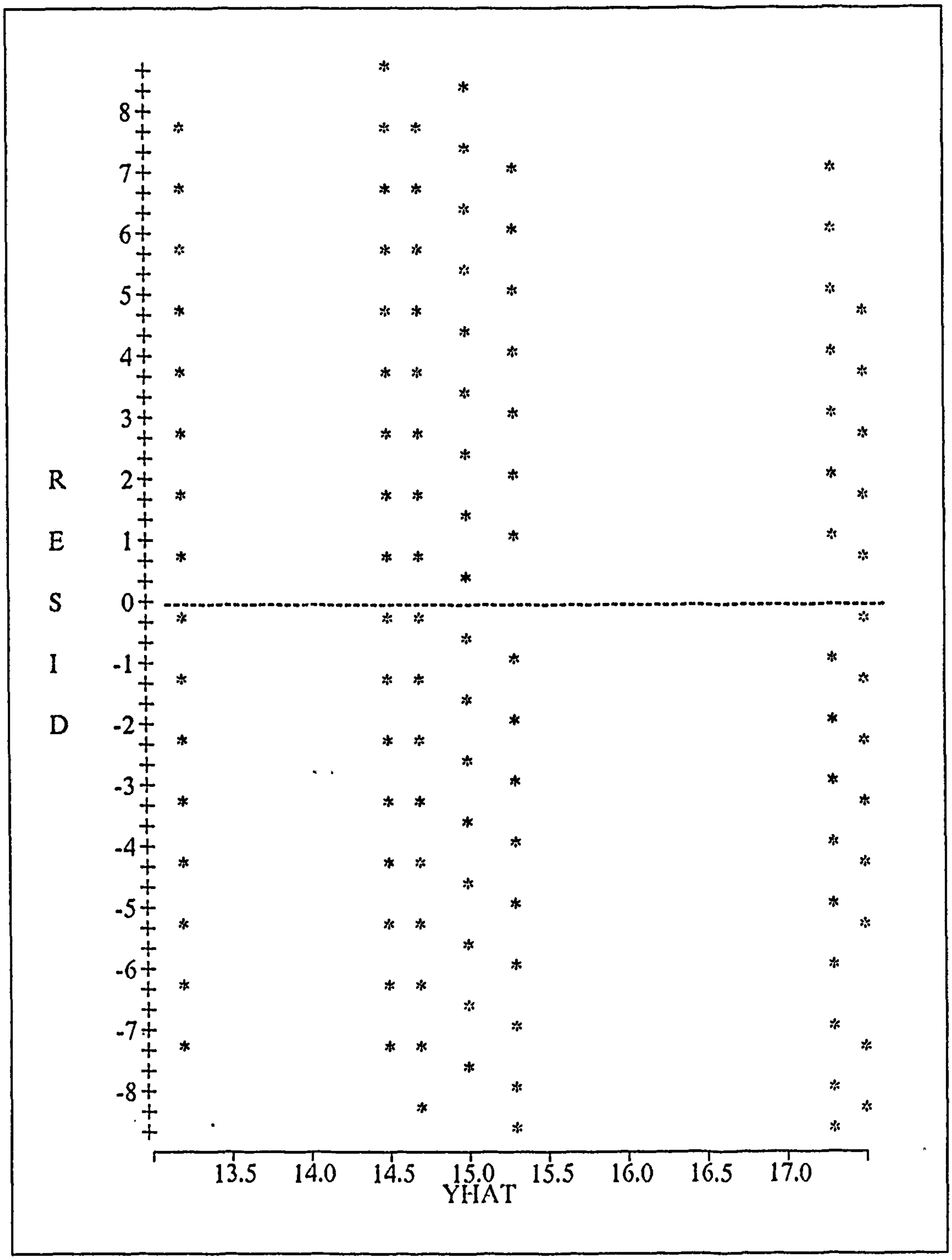


Figure A.7 AW ANOVA: Time to get rated - RESID vs YHAT.

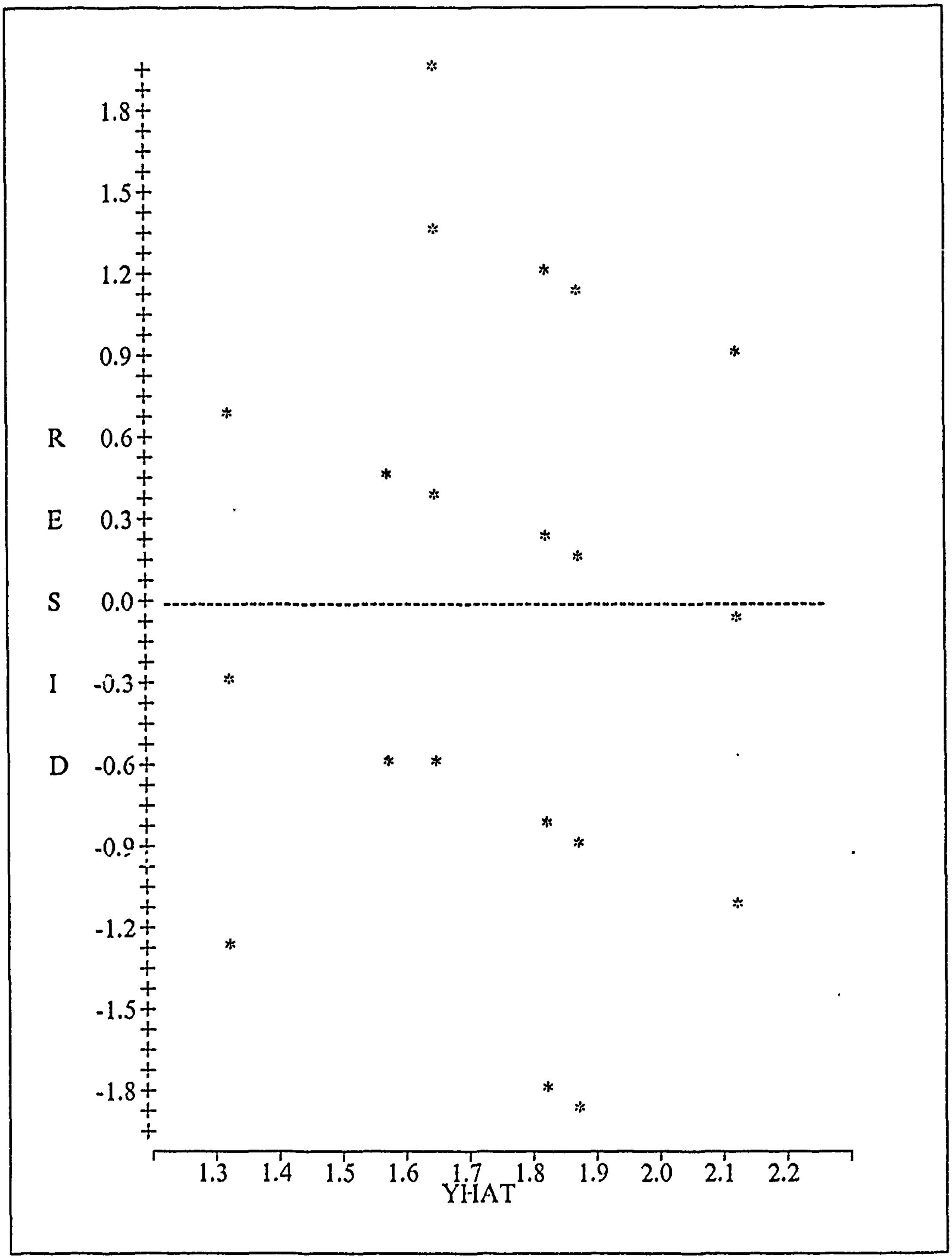


Figure A.8 AW ANOVA: NEC's per year group - RESID vs YHAT.

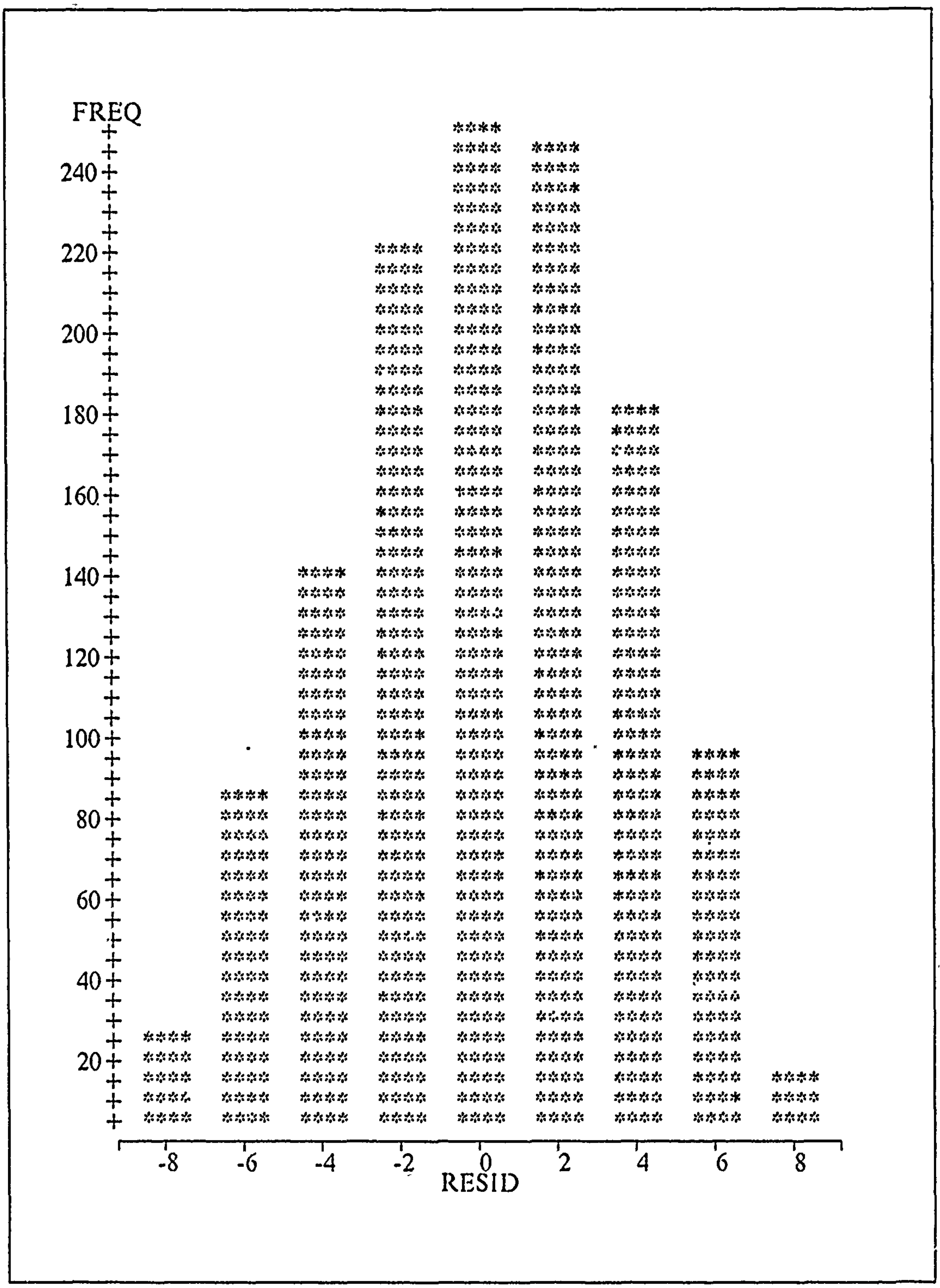


Figure A.9 AW ANOVA: Time to get rated - Histogram of residuals.

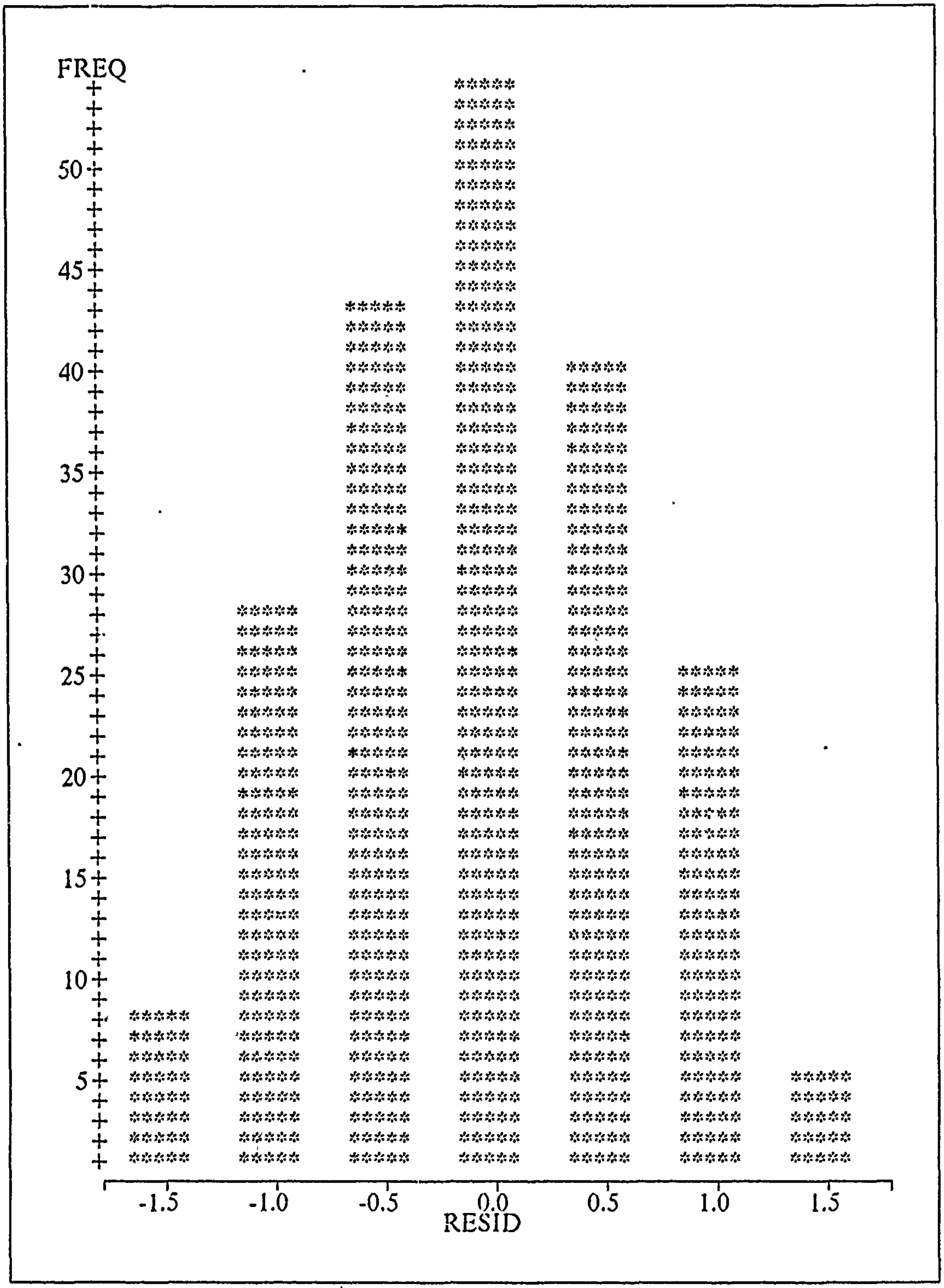


Figure A.10 AW ANOVA: NEC's per year group - Histogram of residuals.

APPENDIX B

DATA BASE

The data base used in this study is described below. Column one is the variable list. Column two gives the location of the variable within the data base. Column three is a description of the variable.

FIELD	POSITION	DESCRIPTION
RECNUM	001-009	Record number
PGMCODE	010-013	Program Code (SG)
RATING	014-016	Rating: AT = Aviation Technician AW = Aviation ASW Operator AX = Aviation ASW Technician
AREA	017-017	Recruiting Area (1, 3, 4, 5, 7, 8)
SEX	018-018	1 = Male 0 = Female
CIVED	019-020	Civilian education: This is the number of years of education completed.
GEDC	021-021	Graduate education code: 1 = High school diploma graduate 2 = Probable Graduate 3 = Graduate equivalence diploma 4 = Non-High school graduate
WAVCE	022-022	Waiver for civilian education: 1 = The recruit received a waiver for entry into the the rating program desired due to lack of sufficient education 0 = otherwise.
AFQT	023-025	Armed Forces Quotient Test score
TESTSW	026-026	Test score waiver: 1 = The recruit received a waiver for entry into the the rating program desired due to low test scores. 0 = otherwise.
GS	027-029	General Science test score
AR	030-032	Arithmetic Reasoning test score
WK	033-035	Word Knowledge test score
PC	036-038	Paragraph Comprehension test score
ND	039-041	Numerical Operations test score
CS	042-044	Coding Speed test score

AS	045-047	Auto Shop test score
MK	048-050	Math Knowledge test score
MC	051-053	Mechanical Comprehension test score
EI	054-056	Electronic Information test score
RACE	057-057	Race: C = Caucasian B = Black X = Other Z = Unknown R = American Indian M = Asian
PAYGRD	058-058	Initial Paygrade (1-9)
SCREEN	059-061	Screen Score: This is the probability a recruit will complete one year of service. Screen scores were developed at CNA.
MATCHS	062-062	Matched SCAT Flag: 1 = yes 0 = no SCAT = System Consolodation for Accessions and Trainees
RTCFLG	063-063	Recruit Training Command Flag: 1 = completed Boot Camp 0 = did not complete Boot Camp
PDEPS	064-064	Primary Dependents: 0 = no primary dependents 1 = spouse only 2 = spouse and 1 child ... 9 = spouse and 8 children or more A = no spouse and 1 child ... H = no spouse and 8 children or more
RATE	065-068	Present Rating
PAYGR	069-069	Present Paygrade
ADSD	070-075	Active Duty Service Date
PEBD	076-079	Pay Entry Base Date
EAOS	080-083	End of Active Obligated Service
COMPLD	084-087	Year-Month NITRAS course completed
COURSE	088-091	NITRAS course code
STUDAC	092-093	NITRAS student action code NITRAS = Navy Integrated Training System
AGE	094-096	Age of recruit
LEFTNAV	097-097	Left Navy flag: 1 = person left the Navy 0 = person did not leave the Navy

ATEAOS	098-098	EAOS Flag: 1 = person left at EAOS 0 = otherwise
MOSIN	099-101	Months in the Navy: Given a person left the Navy, this is the number of months on active duty. If the person is still on active duty, the field is coded '0'.
COMPS	102-104	Composite test score
RATEF	105-108	Final Rating
BLANK	109-109	Blank column
E2	110-112	Months in paygrade E2
E3	113-115	Months in paygrade E3
E4	116-118	Months in paygrade E4
E5	119-121	Months in paygrade E5
INITRAT	122-125	Initial Rating
RATCHG	126-127	Number of Rating Changes
DATE	128-130	Month-Year of change
RATING	131-134	Rating Code
DATE	135-137	Month-Year of change
RATING	138-141	Rating Code
DATE	142-144	Month-Year of change
RATING	145-148	Rating Code
PAYCHG	149-150	Number of Paygrade Changes
DATE	151-154	Year-Month of change
PAYGRADE	155-155	Paygrade Code
DATE	156-159	Year-Month of change
PAYGRADE	160-160	Paygrade Code
DATE	161-164	Year-Month of change
PAYGRADE	165-165	Paygrade Code
DATE	166-169	Year-Month of change
PAYGRADE	170-170	Paygrade Code
DATE	171-174	Year-Month of change
PAYGRADE	175-175	Paygrade Code
NECCHG	176-177	Number of NEC changes
DATE	178-180	Month-Year of change
NEC	181-184	NEC code
DATE	185-187	Month-Year of change

APPENDIX C

PROGRAM LISTING

We built our models using the SAS programming language. SAS provided us numerical computation, statistical results, and graphical summaries. SAS programs used by this study are listed in the following order.

Basic Training:	Two Factor ANOVA Model Single Factor ANOVA Model
Attrition:	Non-Linear Exponential Model Simple Linear Regression Model
Specialized Training:	Simple Linear Regression Model Single Factor ANOVA Model

BASIC TRAINING

Two Factor ANOVA Model

```
OPTIONS LINESIZE=80;
DATA MTGR;
     TITLE 'MONTHS TO GET RATED' ;
     INPUT I M R $ Y;
          LABEL I = ID NUMBER;
          LABEL M = MONTHS TO GET RATED;
          LABEL R = RATING;
          LABEL Y = YEAR;
     CARDS;

PROC GLM  DATA=MTGR / P CLI;
     ID I;
     CLASS R Y;
     MODEL M = R Y R*Y / P CLI;
     MEANS Y;
     MEANS Y/TUKEY;
     OUTPUT OUT = STATS  P = YHAT  R = RESID;
PROC PLOT DATA = STATS;
     PLOT M * Y = '*' ;
```

Single Factor ANOVA Model

```
OPTIONS LINESIZE=80;
DATA BTP;
     TITLE 'MONTHS TO GET RATED' ;
     INPUT I M Y;
          LABEL I = ID NUMBER;
          LABEL M = MONTHS TO GET RATED;
          LABEL Y = YEAR;
     CARDS;

PROC GLM  DATA=BTP;
     ID I;
     CLASS Y;
     MODEL M = Y / P CLI;
     MEANS Y/TUKEY;
     OUTPUT OUT = STATS  P = YHAT  R = RESID;
PROC PLOT DATA = STATS;
     PLOT M * Y;
     PLOT YHAT * Y = '*' ;
     PLOT RESID * Y = '*' / VREF = 0 ;
     PLOT RESID * YHAT = '*' / VREF = 0;
PROC CHART DATA = STATS;
     VBAR RESID;
```

ATTRITION

Non Linear Exponential Model

```
OPTIONS LINESIZE=80;
DATA AR;
    TITLE 'ATTRITION RATE     E(NT) = N*EXP(-LAMBDA*T)' ;
    INPUT T L CL NT P R Y;
    N = XX;
        LABEL T  = TIME IN MONTHS;
        LABEL L  = LOSS;
        LABEL CL = CUMULATIVE LOSS;
        LABEL NT = NUMBER OF SURVIVORS AT TIME T;
        LABEL P  = PERCENT OF SURVIVORS AT TIME T;
        LABEL R  = RATING;
        LABEL Y  = YEAR GROUP;
    CARDS;
;
PROC NLIN DATA=TAR;
    PARAMETERS LAMBDA = .01;
    MODEL NT = N*EXP(-LAMBDA*T);
    DER.LAMBDA = -N*T*EXP(-LAMBDA*T);
    OUTPUT OUT = STATS    P = NTHAT    R = RESID;
PROC PLOT DATA = STATS;
    PLOT NT*T = 'A'    NTHAT*T = 'P' / OVERLAY;
    PLOT RESID * T / VREF = 0;
    PLOT RESID * NTHAT / VREF = 0;
PROC CHART DATA = STATS;
    VBAR RESID;
```

Simple Linear Regresstion Model

```
OPTIONS LINESIZE=80;
DATA AR;
    TITLE 'ATTRITION RATES' ;
    INPUT A R Y X;
        LABEL A = ATTRITION RATE;
        LABEL R = RATING;
        LABEL Y = YEAR GROUP;
        LABEL X = GROUP NUMBER;
    CARDS;
;
PROC REG  DATA=AR;
    MODEL A = X / DW P R CLI CLM INFLUENCE;
    OUTPUT OUT=STATS  P=PRED  R=RESID
           COOKD=CD  H=HAT  RSTUDENT=RS;
PROC PLOT DATA = STATS;
    PLOT A*X;
    PLOT PRED*X;
    PLOT RESID*X / VREF = 0;
    PLOT RESID*PRED / VREF = 0;
    PLOT HAT*X / VREF = 0;
    PLOT RS*A / VREF = 0;
    PLOT CD*X / VREF = 0;
PROC CHART;
    VBAR RESID;
```

Simple Linear Regression Model

```
OPTIONS LINESIZE=80;
DATA NECAVG;
    TITLE 'NECS PER INDIVIDUAL';
    INPUT N S A Y X;
        LABEL N = NUMBER OF NECS;
        LABEL S = SIZE OF YEAR GROUP;
        LABEL A = AVERAGE NUMBER OF NEC'S PER INDIVIDUAL;
        LABEL Y = YEAR GROUP;
        LABEL X = SUBGROUP;
    CARDS;

PROC REG  DATA=TNECAVG;
    ID X;
    MODEL A = X / DW P R CLI CLM INFLUENCE;
    OUTPUT OUT=STATS  P=PRED  R=RESID
           COOKD=CD  H=HAT  RSTUDENT=RS;
PROC PLOT DATA = STATS;
    PLOT A*X;
    PLOT PRED*X;
    PLOT RESID*X / VREF = 0;
    PLOT RESID*PRED / VREF = 0;
    PLOT HAT*X / VREF = 0;
    PLOT RS*A / VREF = 0;
    PLOT CD*X / VREF = 0;
PROC CHART;
    VBAR RESID;
```

Single Factor Anova Model

```
OPTIONS LINESIZE=80;
DATA TNEC;
    TITLE 'NECS PER YEAR GROUP';
    INPUT I M N Y R;
    S = M + N;
        LABEL I = ID NUMBER;
        LABEL M = SECOND YEAR NUMBER OF NECS;
        LABEL N = THIRD YEAR NUMBER OF NECS;
        LABEL Y = YEAR;
        LABEL R = RATING;
    CARDS;

PROC GLM  DATA=TNEC;
    ID I;
    CLASS Y;
    MODEL S = Y / P CLI;
    MEANS Y;
    MEANS Y/TUKEY;
    OUTPUT OUT = STATS  P = SHAT  R = SRESID;
PROC PLOT DATA = STATS;
    PLOT S * Y;
    PLOT SHAT * Y;
    PLOT SRESID * Y / VREF = 0 ;
    PLOT SRESID * SHAT / VREF = 0;
PROC CHART DATA = STATS;
    VBAR SRESID;
```

APPENDIX D

DATA VECTORS

These are the numerical values we used in the SAS programs. Numbers used in the first two ANOVA models will not be listed.

ATTRITION

Non-Linear Regression Data Set

AT	AW	AX
	00 00 00 131 1.000 2 77	
	01 06 06 125 0.954 2 77	
	02 02 08 123 0.938 2 77	
	03 02 10 121 0.923 2 77	
	04 01 11 120 0.916 2 77	
	05 01 12 119 0.908 2 77	
	06 02 14 117 0.893 2 77	
	07 01 15 116 0.885 2 77	
	08 01 16 115 0.877 2 77	
	19 01 17 114 0.870 2 77	
	20 01 18 113 0.862 2 77	
	00 00 00 173 1.000 2 78	
	01 12 12 161 0.930 2 78	
	02 05 17 156 0.901 2 78	
	05 01 18 155 0.895 2 78	
	06 01 19 154 0.890 2 78	
	09 01 20 153 0.884 2 78	
	11 01 21 152 0.878 2 78	
	12 01 22 151 0.872 2 78	
	16 01 23 150 0.867 2 78	
	17 02 25 148 0.855 2 78	
	18 01 26 147 0.849 2 78	
	20 01 27 146 0.843 2 78	
	21 01 28 145 0.838 2 78	
	23 02 30 143 0.826 2 78	
	24 01 31 142 0.820 2 78	
	00 00 00 330 1.000 2 79	
	01 19 19 311 0.942 2 79	
	02 19 38 292 0.884 2 79	
	04 02 40 290 0.878 2 79	
	05 02 42 288 0.872 2 79	
	06 02 44 286 0.866 2 79	
	08 01 45 285 0.863 2 79	
	09 03 48 282 0.854 2 79	
	10 01 49 281 0.851 2 79	
	11 01 50 280 0.848 2 79	
	12 02 52 278 0.842 2 79	
	13 01 53 277 0.839 2 79	
	14 02 55 275 0.833 2 79	
	15 02 57 273 0.827 2 79	
	16 01 58 272 0.824 2 79	
	22 02 60 270 0.818 2 79	
	23 01 61 269 0.815 2 79	
	00 00 00 324 1.000 2 80	
	01 10 10 314 0.969 2 80	
	02 11 21 303 0.935 2 80	

03	02	23	301	0.929	2	80
04	02	25	299	0.922	2	80
05	02	27	297	0.916	2	80
06	01	28	296	0.913	2	80
07	01	29	295	0.910	2	80
08	01	30	294	0.907	2	80
09	03	33	291	0.898	2	80
10	03	36	288	0.888	2	80
11	01	37	287	0.885	2	80
12	03	40	284	0.876	2	80
14	01	41	283	0.873	2	80
15	01	42	282	0.870	2	80
16	01	43	281	0.867	2	80
17	01	44	280	0.864	2	80
21	01	45	279	0.861	2	80
22	02	47	277	0.854	2	80
23	01	48	276	0.851	2	80
24	01	49	275	0.848	2	80

00	00	00	291	1.000	1	81
01	07	07	284	0.975	1	81
02	03	10	281	0.965	1	81
03	01	11	280	0.962	1	81
05	01	12	279	0.958	1	81
11	01	13	278	0.955	1	81
12	02	15	276	0.948	1	81
14	01	16	275	0.945	1	81
15	01	17	274	0.941	1	81
17	02	19	272	0.934	1	81
18	02	21	270	0.927	1	81
19	01	22	269	0.924	1	81
21	01	23	268	0.920	1	81
22	01	24	267	0.917	1	81
24	01	25	266	0.914	1	81

00	00	00	315	1.000	2	81
01	27	27	288	0.914	2	81
02	11	38	277	0.879	2	81
03	04	42	273	0.866	2	81
04	04	46	269	0.853	2	81
05	02	48	267	0.847	2	81
06	01	49	266	0.844	2	81
07	04	53	262	0.831	2	81
08	02	55	260	0.825	2	81
09	02	57	258	0.819	2	81
10	01	58	257	0.815	2	81
13	01	59	256	0.812	2	81
14	02	61	254	0.806	2	81
15	01	62	253	0.803	2	81
17	01	63	252	0.799	2	81
18	01	64	251	0.796	2	81
21	01	65	250	0.793	2	81
22	02	67	248	0.787	2	81
24	01	68	247	0.784	2	81

00	00	00	051	1.000	3	81
01	03	03	048	0.941	3	81
02	02	05	046	0.901	3	81
03	01	06	045	0.882	3	81
06	02	08	043	0.843	3	81
08	01	09	042	0.823	3	81
11	01	10	041	0.803	3	81
16	01	11	040	0.784	3	81
18	01	12	039	0.764	3	81
19	01	13	038	0.745	3	81

00	00	00	664	1.000	1	82
01	15	15	649	0.977	1	82
02	12	27	637	0.959	1	82
03	07	34	630	0.948	1	82
04	01	35	629	0.947	1	82
05	02	37	627	0.944	1	82
06	03	40	624	0.939	1	82
07	02	42	622	0.936	1	82
09	02	44	620	0.933	1	82
10	03	47	617	0.929	1	82
11	04	51	613	0.923	1	82
12	02	53	611	0.920	1	82
13	05	58	606	0.912	1	82
14	02	60	604	0.909	1	82
15	01	61	603	0.908	1	82
16	01	62	602	0.906	1	82
17	01	63	601	0.905	1	82
18	02	65	599	0.902	1	82
19	01	66	598	0.900	1	82
20	03	69	595	0.896	1	82
21	03	72	592	0.891	1	82
22	02	74	590	0.838	1	82
23	02	76	583	0.885	1	82
24	02	78	586	0.882	1	82

00	00	00	501	1.000	2	82
01	45	45	456	0.910	2	82
02	22	67	434	0.866	2	82
03	03	70	431	0.860	2	82
04	03	73	428	0.854	2	82
05	02	75	426	0.850	2	82
07	03	78	423	0.844	2	82
08	02	80	421	0.840	2	82
09	01	81	420	0.838	2	82
10	02	83	418	0.834	2	82
11	01	84	417	0.832	2	82
12	03	87	414	0.826	2	82
13	02	89	412	0.822	2	82
14	01	90	411	0.820	2	82
15	01	91	410	0.818	2	82
19	01	92	409	0.816	2	82
21	01	93	408	0.814	2	82
22	02	95	406	0.810	2	82
23	02	97	404	0.806	2	82
24	01	98	403	0.804	2	82

00	00	00	177	1.000	3	82
01	03	03	174	0.983	3	82
02	02	05	172	0.971	3	82
03	01	06	171	0.966	3	82
07	01	07	170	0.960	3	82
09	01	08	169	0.954	3	82
11	02	10	167	0.943	3	82
14	01	11	166	0.937	3	82
15	02	13	164	0.926	3	82
18	01	14	163	0.920	3	82
19	01	15	162	0.915	3	82
23	01	16	161	0.909	3	82

00	00	00	455	1.000	1	83
01	06	06	449	0.986	1	83
02	09	15	440	0.967	1	83
03	01	16	439	0.964	1	83
05	02	18	437	0.960	1	83

00	00	00	432	1.000	2	83
01	25	25	407	0.942	2	83
02	14	39	393	0.909	2	83
03	05	44	388	0.898	2	83
04	10	54	378	0.875	2	83

00	00	00	099	1.000	3	83
01	03	03	096	0.969	3	83
02	02	05	094	0.949	3	83
03	02	07	092	0.929	3	83
04	01	08	091	0.919	3	83

06 01 19 436 0.958 1 83	05 11 65 367 0.849 2 83	
08 01 20 435 0.956 1 83	06 02 67 365 0.844 2 83	
09 02 22 433 0.951 1 83	07 01 68 364 0.842 2 83	
10 03 25 430 0.945 1 83	08 02 70 362 0.837 2 83	
11 01 26 429 0.942 1 83	09 03 73 359 0.831 2 83	
15 01 27 428 0.940 1 83	10 02 75 357 0.826 2 83	
16 01 28 427 0.938 1 83	11 01 76 356 0.824 2 83	
17 01 29 426 0.936 1 83	12 01 77 355 0.821 2 83	
	13 02 79 353 0.817 2 83	
	14 01 80 352 0.814 2 83	
	15 01 81 351 0.812 2 83	
	16 01 82 350 0.810 2 83	
	17 01 83 349 0.807 2 83	
00 00 00 053 1.000 1 84	00 00 00 053 1.000 2 84	00 00 00 0 9 1.000 3 84
02 01 01 052 0.981 1 84	01 02 02 051 0.962 2 84	01 01 01 0 8 0.888 3 84
06 01 02 051 0.962 1 84	02 02 04 049 0.924 2 84	06 01 02 0 7 0.777 3 84
	08 01 05 048 0.906 2 84	14 01 03 0 6 0.666 3 84
	11 01 06 047 0.887 2 84	

Linear Regression Data Set

AT	AH	AX
.00408278 1 81 1	.01021598 2 77 1	.01747914 3 81 1
.00602720 1 82 2	.00937421 2 78 2	.00470880 3 82 2
.00475758 1 83 3	.01280403 2 79 3	.02313339 3 83 3
.00673668 1 84 4	.00872095 2 80 4	.03262699 3 84 4
	.01447737 2 81 5	
	.01271186 2 82 6	
	.01681385 2 83 7	
	.01230967 2 84 8	

SPECIALIZED TRAINING

Linear Regression Data Set

AT	AH	AX
369 232 1.5905 81 1	114 070 1.6286 77 1	058 033 1.7576 81 1
1010 524 1.9275 82 2	154 102 1.5098 78 2	255 139 1.8345 82 2
619 365 1.6959 83 3	349 165 2.1152 79 3	133 079 1.6835 83 3
	422 213 1.9812 80 4	
	352 177 1.9887 81 5	
	668 304 2.1974 82 6	
	444 243 1.8272 83 7	

Analysis of Variance Data Set

AT	AH	AX
0015 2 1 83 1	0015 0 0 79 2	0002 0 1 81 3
0052 2 1 82 1	0019 1 0 78 2	0003 1 0 83 3
0072 1 0 82 1	0020 1 1 83 2	0005 2 0 82 3
0074 1 1 83 1	0042 1 1 78 2	0006 1 1 82 3
0087 1 0 82 1	0052 0 2 81 2	0017 0 1 83 3
0090 2 0 83 1	0076 2 0 80 2	0020 2 0 82 3
0095 0 1 81 1	0081 0 1 81 2	0022 1 1 82 3
0111 0 1 81 1	0082 2 0 82 2	0025 1 0 83 3
0120 1 1 83 1	0096 1 1 83 2	0026 0 1 82 3
0151 0 1 83 1	0098 1 1 81 2	0032 1 1 83 3
0173 1 0 83 1	0099 1 0 83 2	0033 0 1 81 3

0182	2	0	81	1
0215	1	1	81	1
0232	2	0	82	1
0254	1	0	83	1
0299	1	1	81	1
0331	0	2	81	1
0346	1	0	83	1
0347	2	0	82	1
0353	1	0	83	1
0356	2	2	82	1
0358	0	1	81	1
0377	0	2	81	1
0378	0	2	82	1
0383	2	0	81	1
0384	1	1	82	1
0390	1	1	82	1
0415	0	1	83	1
0424	1	0	83	1
0461	1	1	81	1
0463	0	1	82	1
0501	2	0	83	1
0502	1	0	82	1
0507	1	0	82	1
0541	0	0	81	1
0550	1	0	81	1
0558	2	0	81	1
0579	2	0	83	1
0605	2	0	82	1
0627	2	1	81	1
0662	1	1	81	1
0701	0	2	82	1
0709	1	1	82	1
0718	2	0	83	1
0753	1	0	83	1
0757	1	1	83	1
0768	1	1	83	1
0781	1	1	83	1
0784	1	0	81	1
0794	1	0	83	1
0804	2	0	81	1
0907	0	2	82	1
0908	2	0	83	1
0909	2	1	82	1
0945	1	1	82	1
1043	2	0	83	1
1050	1	1	82	1
1066	2	1	83	1
1070	0	1	83	1
1074	1	1	82	1
1083	1	0	83	1
1095	2	1	82	1
1102	2	0	82	1
1113	1	0	82	1
1120	2	0	82	1
1128	1	0	81	1
1175	1	1	81	1
1220	1	1	83	1
1244	1	1	81	1
1262	2	0	81	1
1285	1	0	81	1
1309	1	2	82	1
1321	1	0	82	1
1325	2	0	83	1
1326	1	1	81	1
1332	1	1	82	1
1333	1	0	83	1
1334	0	1	83	1
1340	1	1	82	1
1345	0	1	81	1
0102	1	1	77	2
0112	2	0	79	2
0121	1	1	82	2
0141	0	1	77	2
0155	0	2	79	2
0161	1	0	83	2
0184	1	1	77	2
0193	0	1	81	2
0197	0	0	78	2
0200	1	0	78	2
0202	2	1	79	2
0215	1	0	83	2
0232	0	1	80	2
0251	0	1	82	2
0254	1	0	78	2
0269	2	0	82	2
0275	1	1	77	2
0276	0	0	79	2
0307	0	2	81	2
0316	1	1	83	2
0370	1	1	77	2
0379	1	2	80	2
0396	1	0	83	2
0399	0	1	80	2
0404	0	2	79	2
0411	2	1	83	2
0427	1	1	82	2
0435	1	2	82	2
0460	2	0	80	2
0461	1	0	83	2
0474	1	1	77	2
0499	2	0	78	2
0512	1	0	77	2
0536	1	0	79	2
0542	1	0	83	2
0547	0	2	79	2
0561	0	0	79	2
0578	1	1	80	2
0593	1	0	78	2
0603	1	1	79	2
0606	0	2	78	2
0616	2	0	80	2
0617	1	1	78	2
0643	0	2	79	2
0663	1	1	81	2
0671	1	0	83	2
0686	1	1	78	2
0688	1	1	81	2
0689	1	0	77	2
0691	2	0	82	2
0709	1	1	78	2
0718	1	0	83	2
0723	1	2	81	2
0751	0	1	79	2
0772	1	0	78	2
0780	1	1	78	2
0791	1	1	82	2
0792	0	2	80	2
0801	2	1	82	2
0817	1	1	77	2
0818	1	0	78	2
0819	1	1	81	2
0822	1	1	81	2
0824	1	1	81	2
0826	1	1	83	2
0842	0	1	77	2
0846	0	1	81	2
0853	1	0	78	2
0888	1	1	80	2
0049	1	1	83	3
0051	1	1	81	3
0054	1	2	81	3
0056	1	0	82	3
0057	0	2	81	3
0060	1	2	81	3
0065	0	2	82	3
0066	2	0	82	3
0067	1	1	82	3
0072	2	0	83	3
0073	1	1	82	3
0079	0	1	81	3
0081	1	0	81	3
0087	2	1	82	3
0088	0	1	83	3
0089	0	1	81	3
0092	0	2	81	3
0093	0	1	82	3
0094	0	1	81	3
0097	2	0	83	3
0101	1	1	81	3
0108	1	1	81	3
0110	1	0	81	3
0111	1	1	82	3
0117	0	1	81	3
0127	1	1	81	3
0138	1	0	82	3
0147	0	2	82	3
0153	0	1	83	3
0162	1	0	82	3
0166	1	0	83	3
0172	1	1	82	3
0173	1	0	83	3
0175	0	2	83	3
0177	0	1	83	3
0182	0	1	81	3
0184	2	0	81	3
0187	0	0	83	3
0192	1	0	82	3
0195	1	0	82	3
0200	1	0	82	3
0203	1	1	81	3
0204	1	1	82	3
0206	1	0	82	3
0208	1	1	81	3
0213	1	2	81	3
0217	1	1	82	3
0221	1	0	82	3
0227	0	1	82	3
0236	1	1	83	3
0237	1	0	82	3
0243	1	0	83	3
0249	1	2	83	3
0251	1	0	83	3
0255	1	0	83	3
0258	1	1	82	3
0259	2	0	83	3
0261	0	2	81	3
0264	1	2	81	3
0269	0	1	81	3
0271	1	0	83	3
0272	2	0	82	3
0274	1	0	83	3
0282	2	0	83	3
0287	2	0	83	3
0290	1	0	83	3
0295	1	0	83	3
0296	0	3	81	3
0300	1	1	83	3

1367	1	0	83	1
1371	1	1	81	1
1418	1	1	81	1
1420	1	2	81	1
1424	0	1	81	1
1440	0	2	82	1
1454	1	1	81	1
1455	2	0	81	1
1459	1	0	82	1

0897	1	1	81	2
0926	1	1	78	2
0949	1	0	83	2
0954	1	1	77	2
0969	1	1	77	2
0971	1	2	80	2
0988	2	0	81	2
0994	1	1	83	2
0998	1	0	83	2
1010	1	0	81	2
1014	0	1	78	2
1018	1	2	82	2
1056	1	0	83	2
1057	1	1	79	2
1062	1	1	82	2
1073	1	0	78	2
1078	1	0	77	2
1079	0	1	77	2
1090	0	2	80	2
1097	0	1	83	2
1108	1	1	79	2
1111	2	0	79	2
1112	2	0	77	2
1113	1	0	77	2
1118	2	1	79	2
1125	0	1	80	2
1132	0	1	79	2
1145	2	0	81	2
1155	1	2	82	2
1168	1	1	80	2
1194	2	1	83	2
1201	0	2	80	2
1220	1	1	82	2
1221	1	1	77	2
1256	2	1	80	2
1258	1	1	83	2
1270	0	0	80	2
1279	1	0	83	2
1285	1	1	79	2
1298	2	1	81	2
1301	1	1	82	2
1304	1	1	82	2
1323	0	1	81	2
1324	2	1	83	2
1349	0	2	81	2
1351	0	1	77	2
1358	1	1	78	2
1386	0	1	77	2
1392	1	0	83	2
1408	0	0	78	2
1415	2	0	78	2
1441	0	2	80	2
1448	1	0	79	2
1462	1	1	81	2
1479	1	1	82	2
1484	0	1	79	2
1497	1	1	81	2
1498	1	0	79	2
1499	1	1	78	2
1513	1	1	77	2
1522	0	2	80	2
1541	0	1	82	2
1552	1	2	80	2
1568	0	1	78	2
1577	2	0	83	2
1598	1	1	80	2
1602	0	2	77	2
1624	1	1	77	2
1630	1	0	78	2

0303	0	2	81	3
0306	1	0	83	3
0307	1	0	81	3
0312	0	2	82	3
0317	0	1	81	3
0322	0	2	82	3
0323	1	1	81	3
0325	0	2	81	3
0333	0	1	83	3

1641	0	1	80	2
1642	1	1	79	2
1652	2	0	80	2
1666	0	1	77	2
1673	1	1	82	2
1693	0	1	83	2
1711	1	2	80	2
1721	1	0	81	2
1729	1	0	80	2
1730	1	0	81	2
1736	1	1	81	2
1749	0	1	80	2
1757	1	1	82	2
1761	1	1	77	2
1763	0	1	81	2
1766	1	0	78	2
1777	1	2	79	2
1782	1	2	82	2
1786	2	0	77	2
1788	2	0	83	2
1804	0	2	82	2
1821	1	1	83	2
1827	1	1	77	2
1831	1	1	82	2
1860	1	2	81	2
1874	2	0	81	2
1879	0	1	78	2
1884	2	0	82	2
1895	1	2	81	2
1908	1	1	77	2
1928	0	1	78	2
1945	2	1	79	2
1970	1	2	79	2
1974	0	2	80	2
1975	0	1	77	2
1978	2	0	82	2
1986	1	0	77	2
1989	1	0	79	2
1992	0	1	77	2
2008	1	2	82	2
2018	1	2	82	2
2047	1	0	80	2
2053	1	1	79	2
2125	0	2	82	2
2163	0	1	82	2
2164	1	2	81	2
2179	2	0	83	2
2181	1	0	78	2
2184	0	0	80	2
2188	1	2	83	2
220001		0	79	2
2204	0	0	78	2
2206	0	3	80	2
2209	2	0	83	2
2215	0	1	81	2
2217	1	0	82	2
2231	1	1	82	2
2241	0	2	78	2
2255	2	1	80	2
2260	2	1	79	2

LIST OF REFERENCES

1. Corliss, George R., *CNA Research Memorandum 86-3 "Non Prior Service Accession Data Set,"* Alexandria, Virginia, Center for Naval Analyses, 1986.

2. Neter, John, Wasserman, William, and Kutner, Michael H., *Applied Linear Statistical Models*, Richard D. Irwin, Inc., 1985.

3. Conover, W.J., *Practical Nonparametric Statistics*, John Wiley and Sons, Inc., 1980.

4. Draper, N.R. and Smith, H., *Applied Regression Analysis*, John Wiley and Sons, Inc., 1981.

5. Judge, George G., Hill, R. Carter, Griffiths, William E., Lütkepohl, Helmut, and Lee, Tsoung-Chao, *Introduction to the Theory and Practice of Econometrics*, John Wiley and Sons, Inc., 1982

BIBLIOGRAPHY

Barr, Donald R. and Zehna, Peter W., *Probability: Modeling Uncertainty*, Addison-Wesley Publishing Company, 1983.

Box, George E.P., Hunter, William G., and Hunter, J. Stuart, *Statistics for Experimenters*, John Wiley and Sons, Inc., 1978.

Chambers, John M., Cleveland, William S., Kleiner, Beat, and Tukey, Paul A., *Graphical Methods for Data Analysis*, Wadsworth International Group, Belmont, California, 1983

Gilman, Leonard and Rose, Allen J., *APL An Interactive Approach*, Third Edition, John Wiley and Sons, Inc., 1984.

Kreitzberg, Charles B. and Shneiderman, Ben, *Fortran Programming*, Second Edition, Harcourt Brace Jovanovich, Inc., 1982.

Mosteller, Frederick and Tukey, John W., *Data Analysis and Regression*, Addison-Wesley Publishing Company, 1977.

Questor, Aline O. and Corliss, George R., *CNA Research Memorandum 86-84 "The Navy Enlisted Classification (NEC's): A History"*, Alexandria, Virginia, Center for Naval Analyses, 1986.

Questor, Aline O., Byrnes, Patricia E., Corliss, George R., Dorsey, Cheryl A., Hill, John L., and Schoeck, Susan M. *CNA Research Memorandum 86-90 "Specialized Training for Enlisted Navy Personnel: A Historical Account"*, Alexandria, Virginia, Center for Naval Analyses, 1986.

Ramsey, James B. and Musgrave, Gerald L., *APL STAT*, Lifetime Learning Publications, 1981.

SAS Institute Inc., *SAS User's Guide Basics*, Version 5 Edition, Cary, North Carolina, 1985.

SAS Institute Inc., *SAS User's Guide Statistics*, Version 5 Edition, Cary, North Carolina, 1985.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginina 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5000	2
3.	Professor Dan C. Boger, Code 55Bo Naval Postgraduate School Monterey, California 93943-5000	1
4.	Professor Robert R. Read, Code 55Re Naval Postgraduate School Monterey, California 93943-5000	1
5.	Center for Naval Analyses 4401 Ford Avenue Post Office Box 16268 Alexandria, Virginia 22302-0268	2
6.	Lieutenant Kappy Aiu 4855-C Nonou Road Kapaa, Kauai, Hawaii 96746	2